University of Derby

Assessing the Credibility of Online Social Network Messages

Oghenefejiro Winnie Makinde

Doctor of Philosophy 2018

TABLE OF CONTENTS

Chapter 1: Introduction

Chapter 2: Literature Review

Chapter 3: Research Methodology

Chapter 4: The experiment

Chapter 5: Results

Chapter 6: Validation

List of Tables

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of Figures

Chapter 5

Chapter 6

Chapter 8

DECLARATION

The research described in this thesis has not been presented or submitted in support of application for another degree/award in any other university or institution of learning.

ACKNOWLEDGEMENT

I am very thankful to God almighty who gave life and good health to me without a price. To Jesus His son for showering me with unlimited favour and mercies to pull through and to the Holy Spirit for been my comforter in trying times; giving me strength daily throughout this process. This research is a success because of you, for this I say Oghene-do (Thank you).

I will like to thank my parents, Chief Williams Makinde (PhD) and Mrs Flora Makinde for providing the financial support, care and emotional assistance all through this program. I am very grateful to you both for believing in me and encouraging me to get to the finish line. I will also like to thank my siblings, Eguono Makinde, Akpervwe Makinde, Timothy Makinde and Ruona Makinde for supporting and entertaining me with laughter whenever I speak with them. To my loving husband Babatunde Bello, I am grateful and thankful that God gave me such a blessing in you. Thank you for putting up with my continued absent mindedness all these years, cooking, cleaning, encouraging, praying and interceding for me at all times. To my big sister: Sister Debola wow, I am thankful to God for bringing you into my life at such a time as this. Your endless phone calls daily/weekly was of immense help to me, providing uplifting words and prayers all the time; thank you ma.

Special thanks goes to the University of Derby for providing an atmosphere of learning and growth intellectually. Lending me a wonderful supervisor in Professor Richard Hill. Professor Hill was continually encouraging and approachable at all times. Dr. Marcello Trovati for your words of encouragement even when you joined towards the end of the research process. I am very grateful to you both.

To my church Pastor and his wife, Pastor David and Deaconess Precious Bankole, thank you Sir/Ma for your prayers and consistency in making sure I was sound spiritually and

mentality every day. To all my church members (Christ Embassy Nottingham) I say a big thank you.

Finally, I would like to thank all my friends and family members for their support and prayers throughout this program at the University of Derby.

Thank you all so much God bless you.

Publication

O. W. Makinde, Social Network Message Credibility: an Agent-Based Modelling approach, 4th International Conference on Emerging Intelligent Data and Web Technologies - Workshop: Second International Workshop on Cybercrimes and Emerging Web Environments (CEWE), 2013.

ABSTRACT

Information gathered socially online is a key feature of the growth and development of modern society. Presently the Internet is a platform for the distribution of data. Millions of people use Online Social Networks daily as a tool to get updated with social, political, educational or other occurrences. In many cases information derived from an Online Social Network is acted upon and often shared with other networks, without further assessments or judgments. Many people do not check to see if the information shared is credible. A user may trust the information generated by a close friend without questioning its credibility, in contrast to a message generated by an unknown user. This work considers the concept of credibility in the wider sense, by proposing whether a user can trust the service provider or even the information itself. Two key components of credibility have been explored; trustworthiness and expertise.

Credibility has been researched in the past using Twitter as a validation tool. The research was focused on automatic methods of assessing the credibility of sets of tweets using analysis of microblog postings related to trending topics to determine the credibility of tweets.

This research develops a framework that can assist the assessment of the credibility of messages in Online Social Networks. Four types of credibility are explored (experienced, surface, reputed and presumed credibility) resulting in a credibility hierarchy. To determine the credibility of messages generated and distributed in Online Social Networks, a virtual network is created, which attributes nodes with individual views to generate messages in the network at random, recording data from a network and analysing the data based on the behaviour exhibited by agents (an agent-based modelling approach). The factors considered for the experiment design included; peer-to-peer networking, collaboration, opinion formation and network rewiring. The behaviour of agents, frequency in which messages are

shared and used, the pathway of the messages and how this affects credibility of messages is also considered. A framework is designed and the resulting data are tested using the design.

The resulting data generated validated the framework in part, supporting an approach whereby the concept of tagging the message status assists the understanding and application of the credibility hierarchy.

Validation was carried out with Twitter data acquired through twitter's Application Programming Interface (API). There were similarities in the generation and frequency of the message distributions in the network; these findings were also recorded and analysed using the framework proposed. Some limitations were encountered while acquiring data from Twitter, however, there was sufficient evidence of correlation between the simulated and real social network datasets to indicate the validity of the framework.

Chapter 1

Introduction

1.0 Research Background

Social Information is a key feature of the growth and development of modern society. Presently the Internet is a platform for the distribution of data. The Internet is a collection/conglomeration of networks, where a vast amount of data is generated, shared and processed in networks of connected entities.

Since the creation of the World Wide Web, a lot of changes have occurred. Today, online social networks continue to materialise as a significant part of the Web. At present, the term "Social Network" is generally associated with Internet based communities such as MySpace, Facebook, LinkedIn, Twitter and many others (Menges *et al*, 2008). These are virtual communities where participants are able to communicate and distribute information about other members or events that occur. A single individual, the community or the public at large may use the information that is generated.

Millions of people use online social networks daily as a tool to keep up to date with social, political, educational or other occurrences. Amongst those that use these networks are some users who access the Internet to search for information that answer their everyday questions.

In many cases information derived from online social networks is acted upon and often shared with other networks, without any kind of assessment or evaluation. Many people do not verify if the information shared is trustworthy. Just because a friend has uploaded a message online does not mean the message can be trusted. A user is more likely to trust the information generated by a close friend without questioning the credibility of the information as compared to one generated by an unknown user. Trust and credibility between members of a particular group can emerge as more and more information is generated, propagated and acted upon. According to Guimarães *et al* (1994), a reputation mechanism does not always lead to credibility. This was evident in the apology given to the clothing store Primark by the BBC over false footage of child labour being used at a factory in Bangalore. Many who saw the documentary went online to air their views and as a result misled other Internet users (BBC Webpage, 2011).

The concept of credibility has been considered in the wider sense, by proposing whether a user can trust the service provider or even the information itself. Earlier in 1999, Fogg and Tseng discussed some elements of computer credibility. They highlighted two key components of credibility: trustworthiness and expertise. Trustworthiness is defined as been truthful, unbiased, and well intentioned. This component captures the perceived goodness or morality of the source (Fogg and Tseng, 1999). Expertise is defined as being knowledgeable, experienced and competent. This component represents the perceived knowledge and skill of the source (Fogg and Tseng, 1999).

Some assessment of the credibility of a particular online social network might be useful to introduce the possibility of a transparent reputation mechanism. When the background of users in a network is made available to every user in the network, this process could bring some level of transparency and trust to a network. We already see this in the Social Network LinkedIn (www.linkedin.com): users then have to use this information provided to make decisions when interacting with the network. Although there may be some less trustworthy user profiles in LinkedIn, users are given the opportunity to make their decisions based on the available information given by the alleged owner of the profile.

Whilst a scientist may question the source of the information and its intentions, a vast proportion of information is distributed based on the perceived reputation of the author or propagator. Information taken from a well-respected author, blog or company website will most likely not be questioned before it is distributed. As a result of this many users are armed with misleading information without their knowledge and hence distribute this in their network. It is therefore important to find ways of determining when information in Online Social Networks is indeed trustworthy or deemed credible.

1.1 Related Work

Castillo et al, (2011) using twitter as a case study, have studied the subject of Information Credibility. In their work they focused on automatic methods of assessing the credibility of a given set of tweets. The analysis here was on microblog postings related to trending topics to classify tweets as either credible or not credible. The procedure used in this case analysed human assessments about credibility of items on a sample of twitter postings.

These researchers requested the participation of humans in a survey style kind of data collection to indicate what they think a message posted could be, giving 4 options; *almost certainly true*, *likely to be false*, *almost certainly false* and *I can't decide*.

In this research work however, we have created a model to represent what a network operates like without outside interference. We also consider the four types of credibility highlighted by Fogg and Tseng in 1999, aiming to annotate agents with their credibility ranking. We used agent-based modelling in this research process, watching carefully the interactions of agents and the possible effect upon their behaviour, including whether an agent could alter messages or decision making within the network.

1.2 Research Question

In other research work carried out relating to social network messages the focus has been on the source of the information. In Tanaka *et al*, (2010) research done in evaluating the credibility of Web Information states that;

> *"The abundance of content on the Web, the Lack of publishing barriers, and poor quality control of Web content raise credibility issues"*.

They gave an example of the presence of over twenty thousand health-related sites on the Web, where a medical specialist has reviewed less than half. There are also many Blogs/websites giving advice on varied topics to users, but just because a webpage looks 'genuine' does not mean that all of the information uploaded on there is trustworthy.

Online Social Networks are made of communities of users. In our model, each user is said to be a node or an agent. They could either function as individuals or as a group depending on the goal they aim to achieve. The messages circulated on social networks are generated and distributed by individual users (agents) or groups of agents, and are used by many other agents within and outside of their network. An agent will share a message received from another agent, based on the perceived trust that the message is coming from a reliable or credible source. This trust is achieved based on continual interactions between agents. However, this does not guarantee the credibility of the message. Just because Tony has not given false information in the past does not mean he is likely to provide credible guidance in the future. To aid the validation process for many of such messages this research focuses on the behaviour of agents in a network, looking at the frequency of messages shared with other agents and how trust is built to achieve credible communication.

Agent based modelling (ABM) is used to recreate a complex social network, predicting their characteristics and processes. Determining agent behaviour in a network is one of many instances that ABM has been used in. Research carried out in the Philippines used agent based modelling as the tool to analyse land use dynamics in response to farmer behaviour and environmental change in the Pampamga delta (Mialhe *et al*, 2012). Agent based modelling was also used to examine the epistemic consequences of herding behaviour and its surprising robustness (Weisberg, 2013). This research uses ABM to address the question: Can agent-based modelling be a useful indication for the credibility of messages in Online Social Networks?

1.3 Research Aim and Objectives

The aim of this research is to develop a framework that can assist the assessment of the credibility of messages in Online Social Networks (OSN). This will be achieved by considering the following objectives:

1. Critically review the function of Online Social Networks, their agents (users) and how they operate;
2. Evaluate the characteristics of agents and the credibility of the messages they generate;
3. Propose a draft framework to assess the behaviour of agents within an Online Social Network;
4. Evaluate the credibility of message passing between agents;
5. Evaluate the draft framework and identify enhancements and refinements;
6. Discuss and evaluate the framework with some recommendations for implementation.

1.4 The Experiment

To determine the credibility of messages distributed in Online Social Networks, we have generated messages in a simulated network, recorded data from the network and analysed the data based on the behaviour exhibited by agents.

The data analysis was used to understand the following concepts;

a. The behaviour of agents towards other agents; how they communicate and exchange messages.
b. The frequency with which messages are shared and used; the number of times a message is shared or rejected.
c. The pattern/thread of shared messages and how this may affect the credibility of the messages.

1.5 Structure of the thesis

There are seven chapters in this thesis as follows:

1. Chapter 1 presents the research aim and objectives and the importance of this study.
2. In Chapter 2 (Literature review) we evaluate the concept of an Online Social Network, agents and their characteristics, and the concept of trust and credibility for social networks. An introduction to credibility types is provided as well as a hierarchy proposed for the credibility types for Online Social Networks.
3. We present the methods adopted for the research in chapter 3. We evaluate the concept of Agent-based Modelling and how it is been used in other research to achieve various results. We also discuss in detail the specifications and design concept of our virtual network and our expectations of what the data generated during the experiment should illustrate.

4. Chapter 4 describes the various processes carried out in the simulation and how the researcher designed a functional simulation. The approach used for data generation and collection from the simulation based on the credibility hierarchy proposed in chapter one is described. A detailed discussion of how the experiment is expected to run showing how data was collected. This is concluded by a description of each run of the simulation.
5. In Chapter 5 a record of the initial data is documented from the simulation, followed by a set of experiments to achieve the anticipated outcome described in Chapter 4. We also carry out some preliminary analysis to validate the network, processing the scale free characteristics of a real world scenario, as was the intention of the design.
6. Chapter 6 considers the validation and verification of the results with real data from an Online Social network (in this case twitter data). The graphs are evaluated, comparing agent characteristics to the various credibility types.
7. In Chapter 7 we make the concluding arguments, recommendations and propose future work.

Chapter 2

Literature Review

2.0 Introduction

Information distribution is bigger, better and can be shared more quickly with the advent of computers and computer related tools. In the past, according to Fogg and Tseng (1999), the main focus of research was on the credibility of computer systems. This was because many aspects of society began to rely on computers that at the time, were perceived by the general public as machines incapable of making mistakes (Fogg and Tseng, 1999). Credibility became an important factor because users wanted to know how trustworthy computers were and what expertise they had to deliver as well as the quality of service expected from them. Fogg and Tseng, in 1999, discussed the elements of computer credibility, which are:

1. Four types of credibility, specifically the proposal of a conceptual framework of four kinds of credibility.
    - Presumed credibility: This is the presumption that most people tell the truth hence perceiving information is trustworthy.
    - Reputed credibility: Believing the opinion of third parties over a particular topic.
    - Surface credibility: The trust in information based on simple inspection, as in the case of 'judging a book by its cover'.
    - Experienced credibility: Believing in information based on first-hand experience.
2. Credibility evaluation errors: This conceptual framework makes reference to the four types of credibility above, in that users can make evaluations. Such evaluations can either be appropriate or inaccurate. According to Lee (1999), Muir (1987 and

Sheridan (1983), Two kinds of errors are mostly made in making such decisions of trust.

- The first type of error is *Gullibility Error*. For example, in educational institutions, students often make the choice of trusting inaccurate information from a web page even when the web page is not trustworthy. Hence the reason they are taught how to thoroughly evaluate a range of sources of information before referencing them to substantiate an academic argument.
- The second type of error made is called an *Incredibility Error*. This is when users perceive computers to be incredible even when they are credible. This type of error (Fogg and Tseng, 1999) is of great concern to those that design, develop and evaluate computer products.

3. Models for evaluating credibility: Credibility is not only assessed by acceptance or rejection of information as the two errors in the second element may suggest. Rather, as Fogg and Tseng (1999) proposed, there are three prototypical models that may be considered when evaluating credibility. They are:

- Binary evaluation; the product is perceived to be either credible or not credible. This happens when users have:

  > "*low interest in an issue, low ability to process information either due to cognitive abilities or situational factors, little familiarity with the subject matter, [or] no referent point for comparison*" (Fogg and Tseng 1999).

  Any of these factors may lead to a binary decision. A good example is when a student submits an assignment using information from a web page without any kind of validation given to the source.

- Threshold evaluation: This strategy involves an upper or lower credibility assessment. Specifically, when a computer is at the upper threshold (or the threshold is exceeded) it is deemed credible by users. However, when it falls below a lower threshold, it is deemed to be incredible. When the product falls between the two thresholds, then the user may describe the products as having some degree of credibility. Factors that affect this decision include:

> "*Moderate interest in the issue, moderate ability to process information either due to cognitive abilities or situational factors, partial familiarity with the subject matter and moderate ability to compare various sources*" (Fogg and Tseng, 1999).

- Spectral evaluation; this model is the most sophisticated type of evaluation, proposing that users have a spectral strategy to make their decisions on the credibility of a product.

> "High interest in the issue, High ability to process information, including favourable cognitive and situational factors, high familiarity with the subject matter and considerate opportunity to compare various sources". (Fogg and Tseng, 1999).

All of the above elements must be present to facilitate spectral evaluation.

This research analyses the element related to evaluating the credibility of Online Social Network messages, namely the four types of credibility listed in the first element. An evaluation of how agent behaviour through modelling may be classified into each credibility type based on the characteristics portrayed by agents. This is because Online Social Networks have now become the focus of credibility in investigating the trustworthiness of messages generated online (Kim *et al,* 2008).

As human interaction continues to evolve, the use of Internet services have made it easier to communicate with individuals around the world. This is facilitated by tools such as; emails, instant messages and direct messages (currently via Online Social Networks).

2.1 Objectives of this chapter

This chapter includes the following:

1. Introduction to the concept of agents; what agents are and what characteristics they have that makes them function the way they do.
2. Online Social Networks and their functionality; what a network entails and some methods used in analysing these kinds of networks.
3. An evaluation of online social networks highlighting the main area of focus for this research work; the kind of activities that take place together with user activity statistics recorded in recent times.
4. Discussion of opinion formation, trust and credibility in online social networks.
5. The importance of online social network roles as a result of agent collaboration made possible by rewiring within social networks.

It also looks at the challenges of opinion formation as it affects messages distributed in online social networks.

2.2 Agents

An agent or agency is an individual's ability to make free and independent decisions. This is free from any kind of influence or manipulating power by other agents. According to Wooldridge (1995) and Bradshaw *et al*, (1997), an agent is:

> "*an entity that functions continuously and autonomously in a particular environment that is often inhabited by other agents and processes, with agents being autonomous, reactive, proactive and sociable*".

Agents are sometimes referred to as 'software agents' as it is anticipated that they will learn from their experiences, and interconnect and cooperate with other agents within their network (Ismail and Ahmad, 2011). As participating members of a social network, they form edges (connections) with other agents in a network, to achieve both private and public goals. When it is a private task, agents interact with their neighbours, gathering the information required to achieve the result needed. Where the goal is public or group-oriented, agents work together to accomplish said task. In broadcasting, for example, journalists (agents) in an organisation work together searching for evidence of a story and deliberate on how best to report it to the general public. Whether a network is *static* or *dynamic* is determined by the agent's functionality. When new edges are created, there is a dynamic network, but in cases where no new edges are created, there is a static network (Zhang and Leezer, 2009). To achieve a dynamic network, agents practise the act of *coalition* formation. In this system, the agent forming the coalition considers the information about the local neighbourhood before picking new agents to join its network (Barton and Allan, 2008).

All agents in a network are responsible for organising and re-organising the network. An agent-organised network (AON) is:

> "*a set of inter-connected agents who collectively manage the structure of these network agents by making individual decisions about agent to connect based on local information*" [Barton and Allan, 2007 & 2008].

This process is called rewiring.

2.2.1 Rewiring

This is a process of establishing active agents (assets) in a team and which agents are to be taken off the team because they are a liability. An agent's capability, according to Barton and Allan (Barton, 2007), include:

> "*Joining an existing team, initiating a new team, waiting, or rewiring the connections*".

Events in a network can influence an agent's behaviour or opinion when deciding what new agents to take on as part of the team (Kaiser, Krockel and Bodendorf, 2010) but where rewiring is successful, all the agents become active and join in to complete the task in the network. New tasks enter the system over time, which may require rewiring the network because members in the team may not have the skills needed to perform the task. For example, an organisation that produces movies in genre such as drama, will have to hire a professional in animation design when they choose to produce an animation movie.

Rewiring can either cause the addition of credible agents to a network or non-credible agents. Regardless of addition or removal, however, the process is needed for generation of new ideas in a network.

2.2.2 Characteristics of network agents

Software Agents are computational entities that act on behalf of another entity, or entities to perform a task or achieve a given goal (Talia, 2011). They are expected to have the ability to behave with a specific degree of independence, carrying out actions needed to achieve their goals. The agent prototype is designed as a distributed computing model, having the sole objective of interacting with other agents, during which information is exchanged. When this happens collaboration takes place, which enhances the achievement of complex tasks. Nevertheless, a single agent can still perform a given task without the contributions of other agents.

There are several definitions of agents that have been quoted by many authors across different literatures. Some of these include:

- "*An encapsulated computer system that is situated in some environment and that is capable of flexible action in that environment in order to meet its design objectives*" (Jennings, 2001).
- "*Programs that engage in dialogs/negotiations and coordinate the transfer of information*" (Coen, 1991).
- "*Anything that can be viewed as perceiving its environment through sensors and acting upon that environment through effectors*" (Russel and Norvig, 1995).
- "*Autonomous system situated in a dynamical environment acting independently of its restrictions and fulfilling in it a set of goals or directives for which it was created*" (Maes, 1994).

According to Paprzycki and Abraham (2003), there are many different definitions of software agents but whereas some similarities exist between them, they do not seem to define the same entity and can be substituted for each other. There are, however, a wide range of features used to describe a software agent: re-activity, ability to communicate, capacity for cooperation, negotiation, capacity for reasoning, capacity for learning, ability to adapt, intelligence, goal oriented, pro-activity, mobility, robustness, reliability, scalability, flexibility, re-usability, sociability. Apart from these listed features, described by Ismail and Ahmad (2011) as weak notions, there are also features of strong notions, which originate from the humanistic concepts, supporting the mental attributes of agents. Examples include having beliefs, desires, intentions, diligence, knowledge, obligations and commitments.

Some of the features of software agents are explained below (Talia, 2011):

- *Autonomy*: Capacity to act autonomously to some degree on behalf of users or other programs by modifying the way in which they achieve their objectives.
- *Pro-activity*: Ability to pursue individual set goals, including making decisions, which can be influenced by internal decisions.
- *Re-activity*: The capability to react to external events and stimuli, therefore adjust their behaviour and make decisions to carry out their tasks.
- *Communication and Cooperation*: Ability to interact and communicate with other agents (in multiple agent systems), to exchange information, receiving instructions and give responses and cooperate to fulfil their own goals.
- *Negotiation*: Competence in carrying out organized conversations to accomplish a degree of cooperation with other agents.
- *Learning*: Ability to improve performance and decision making over time when interacting with the external environment.

In assessing credibility, an agent must have all the aforementioned characteristics, as each is unique in what it is able or not able to do accordingly. The reliability of an agent is very important to its credibility status but without being intelligent other agents could question its reliability all together.

2.3 Social Network Model

Social networks are popular platforms for communication, interaction and information distribution via the Internet (Wilson *et al*, 2009). The interaction between nodes can be represented by a network or graph. A user is illustrated as a node in a network while the connection between two nodes is referred to as an edge. The presence of a number of edges connected to one node (user) indicates a stronger relation (Ryan, 2011).

This can either be because the node is seen as trustworthy or a link to other nodes, hence multiple connections.

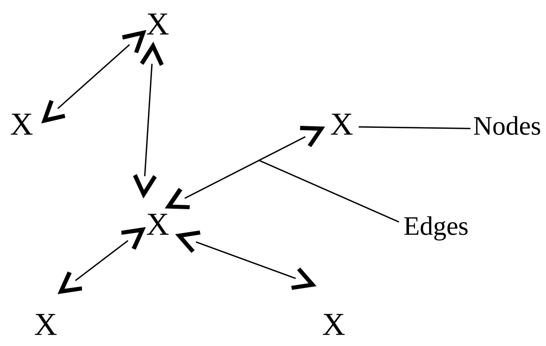

Figure 2.1: Network sketch showing nodes (X) and connecting edges.

A relation is the collection of edges of a specific kind among members of a group (Wasserman, 1994). One scenario is that a node may aim to develop a number of relationships to influence more of the network. By developing a number of relationships with different nodes a user may be able to influence a greater proportion of the network.

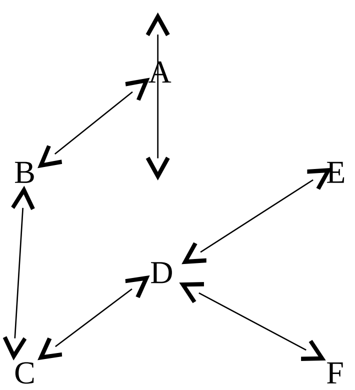

Figure 2.2: Network showing the relationship between nodes (sociogram)

A sociogram is a chart showing the relationship between nodes in a network. It aids in the understanding of a network structure by showing the pattern of connections between nodes (Sherman, 2003). In Figure 2:2 there are six nodes from A to F. Node D in this example is more connected to all users than node B as the figure illustrates. Therefore D will have more information about all six users.

There are three distinctive features in a social network: the agents in the network, the interaction between agents and the information shared (Heidemann, Klier and Probst, 2012).

1. The agents in the Network (these are the nodes in a network). They work together to complete the goal of the network through data generation and sharing.
2. Interaction between agents occurs through communication in the network.
3. Information sharing is information that is shared. This process lets agents achieve various tasks or individual goals as they share or receive information in a network.

A social network is fully functional when all three features take place. For example, in the following twitter conversation:

Jack: “road closure on the M1, road maintenance”

Joel: “Pls what junctions are closed”

Jack: “Btw 22 and 25”

Jack has informed users connected to him of a road closure. Joel received the message but wishes to know which junctions so Jack then replies with an answer. The two agents communicating in this scenario bring about an interaction.

In many social networks a registration process is required from all who wish to become a member, thus creating user profiles for the members. To be a part of Facebook, for example, a user needs to sign up to the platform by completing a form with the following details:

- First name
- Surname
- Email or mobile number
- Re-entering email or mobile number
- New password
- Birthday
- Gender (Male/Female)

Once all these are completed, an account can be created. Other details are required after signing in, such as, for example, location or religion. Once logged in, users can then send requests to friends or family members whom they know are on the network in order to enable broader connections. Over time they will build a bigger network of friends, family and acquaintances.

The moment a profile has been created each member communicates in the network through that profile (network identity), making each user recognizable during a discussion. Twitter and Facebook are some of the many platforms available for interaction within an online social network, Facebook being a popular one (Ellison, Steinfield and Lampe, 2007). These platforms are known to bring people together irrespective of their geographical locations. As Reaney (2012) states:

> "*Most of the world is interconnected thanks to email and social networking sites such as Facebook and Twitter*".

The use of social networks as a tool to discover information now a trend for users of the World Wide Web, these agents, use a combined knowledge of others in a network to answer questions. According to Atzori *et al* 2012, this is the use of collective intelligence to resolve complicated queries. Several nodes communicate to find answers to particular requests made by an individual node. For example, in an online social network, if node 42 were looking for a car rental company with a good reputation it would send this as a post (request) on the network, expecting responses from other nodes. Responses would then be sent by sharing the information on the network, on twitter, this is a re-tweet of the question with suggestions of the answers. The nodes that respond can either be experts or know very little about car rental, as this is not a prerequisite because node 42 made it an open question. Multiple answers are sent many times and node 42 is then left with the task of deciding which answer to go with. It decides its answer based on opinions it already has or on the opinions of its neighbouring nodes.

2.3.1 SOCIAL NETWORK ANALYSIS

Social Network Analysis (SNA) measures relationships between agents in a network through the interactions that take place. This is carried out through the evaluation of behavioural patterns of individual agents. In essence, it investigates how the nodes of the network inter-relate with each other (Furht, 2010). Krebs (2011) has defined SNA as:

> "*the mapping and measuring of relationships and flows between people, groups, organizations, computers, URLs, and other connected information/knowledge entities*".

Mapping a relationship means determining the kind of bond that exists between nodes in a network.

To analyse a network, data from both communication patterns and the frequency of such communications can be observed. This enables the data to be reflected upon in two ways, both of which relate to this research work: ego network analysis and complete network analysis (Borgatti, 1998).

1. Ego network analysis is a survey that shows an individual's relationship based on people they interact with. It answers certain questions on agent characteristics and how these characteristics can affect their communication patterns. These agents do not have to come from a certain group or community.
2. Complete data analysis, on the other hand, is the practice of taking a survey of all the relationships among nodes that are members of the same community. Data from this survey is then analysed comparing node behaviour with that of other nodes in the group.

2.3.2 Methods available to social network analysis (SNA)

There are two methods for analysing social networks: social sciences and mathematical sciences (Pinheiro, 2011). The survey and questionnaire technique is a method from social sciences. It is structured in a flexible manner, enabling the gathering of data from different individuals at the same time (McNabb, 2010). Using questionnaires in a survey gives a researcher the opportunity to decide on which relationship to measure (Carrington, Scott and Wasserman, 2005). This technique can be complex as it is not always definitive which individual or group of individuals will be considered in the collection of data (Knoke and Kuklinski, 1982). Alternatively, the mathematical approach focuses on a visualisation of the data collected.

According to Scott (1991), a graph is used to:

"*describe the pattern of connections amongst points*".

The points refer to nodes; hence data collected is shown in a graph to give a visualisation of the network analysis. The choice of the method used in any research is a matter of preference for the researcher. Ultimately the researcher has to determine which one best suits the research area based on the anticipated result. This is significant as it is dependent on the topic being analysed or the availability of the people needed in terms of data collection. For example, will the research be conducting interviews with people face to face or carrying out online surveys (Qualitative or Quantitative)?.

2.3.3 Online social networks

Online Social Network (OSN) has become one of today's powerful means of communication. Data creation and dissemination is less complex compared to what it was a decade ago, and they have been used for both private and business use. Public relations professionals, social scientists, marketers and researchers continue to find improved ways in which an online platform can or could be used for the distribution of information (Graham, 2010). The information generated is very often propagated into other networks, often without any kind of rigorous assessment or judgment. The phrase "social media" is a term given to a web-based platform for social interaction.

Using web-based and mobile technologies it is possible to turn communication into an interactive dialogue and is referred to as consumer-generated media (CGM) (Jakkola *et al*, 2011). Interaction happens both in the physical world or the virtual world of the Internet. Knowledge, communication, experience and feelings in open and closed societies are shared via social networking; a service provided by social media. Some examples of physical media include newspapers, books, magazines, flyers and billboards. Internet media, on the other hand, can include company websites, personal blogs and interactive discussion websites. In the world today, it is difficult to keep track of the ever-growing business of social media, which is affecting how information generated through this source is validated. Online social media comes in many forms: social networking sites, multimedia sharing, blogs, social marketing sites, micro-blogs and social news. There is an acknowledgement that the method of social media contribution brings about a large audience of followers, which can, in turn, influence society (Treepuech, 2011). Hence the question: ‘can agent-based modelling (which is agent behaviour) be a useful indicator for assessing the credibility of messages in OSNs’?

The availability of these online systems enables people to work together thus encouraging forms of communication and interaction amongst individuals all over the world. Researchers have stated that there are more than 900 social networking sites in existence on the Internet (Bindra *et al*, 2012), with Twitter, Facebook and Google Plus among the most popular (Treepuech, 2011). Statistics from these sites have shown that there is an increased number of people using the Internet for information generation, distribution and consumption (Treepuech, 2011).

- LinkedIn has more than 467 million members worldwide in over 200 countries and territories of which, currently, over 40 million are students and recent college graduates. This online social network operates the world’s largest professional network on the Internet with 106 active users (LinkedIn, 2017).
- Twitter has 320 million monthly active users with over 80% active users on mobile devices (Twitter, 2016).

- Facebook had 1.32 billion active users daily, on average, for June 2017, an increase of 17% yearly. Daily active mobile users 1.15 billion average an increase of 23% yearly (Investor.fb.com, 2017).
- YouTube has over 1 billion users; one-third of the world's population. Hundreds of millions of people watch hours of videos daily, generating a billion views. The average viewing session on mobile devices is now more than 40 minutes (YouTube, 2017).

These figures change constantly (by the second) but show how popular and fast growing the online social media environment has become.

There are certain motivations that attract agents to social networks and for individual agents there are some characteristics they should possess to attain a level of credibility. These motivations, according to (Treepuech, 2011), are outlined thus: anticipated reciprocity, increased recognition, sense of efficacy and sense of community. Research carried out by Vorakulpipat *et al* (2011) on the use of social networking sites in developing countries, focused on Thailand and the UAE. In their study, questionnaires were handed out randomly to 122 members of different social networking sites. The results illustrated that 75% of the respondents said they used the sites because they sought up-to-date information from their community of relatives, friends and acquaintances. Another 67% said that the networks provided them with a platform to write, comment and share information including picture and video uploads. It is easy to accept that connecting with friends and information sharing was the reason why a social networking site like Facebook was created but, nonetheless, the credibility of the information generated cannot be guaranteed. In the same research carried out by Vorakulpipat *et al*, 2011, reported that 15% of the respondents agreed to intentionally posting incorrect or fake information due to varied factors such as been the first to break a news or just to be malicious. They did this by owning multiple accounts a situation that we anticipate can generate false information. Although the reason for having multiple accounts according to 28% of the respondent is an unwillingness to share some personal information with the general public. This then becomes a case of ‘the trust factor’ because if the identity of one particular agent cannot be verified then how can other agents trust the contribution it gives.

Inaccurate information in many instances forms the basis for rumour propagation. It is said that rumoured information spreads rapidly, therefore causing an adverse effect upon the behaviour of people. As a node, the presence of rumours in a network could alter the process and accuracy of completing a task, or even the abandonment of a task altogether. According to Hashimoto *et al*, (2011) in Japan, after the Japanese earthquake and tsunami of 2011, there were rumours propagated using social media concerning the short supply of products like gasoline and toilet roll. Such rumours spread quickly causing chaos and anxiety for the people in the community. The aftermath caused by the subsequent nuclear plant incident resulted in inaccurate accounts being shared via social networks that were not true, causing panic amongst members of the community (Hashimoto *et al*, 2011).

The spread of such (rumoured) messages via online social networks shows that an agent’s behaviour can be altered according to the messages it comes across or receives in a network.

Therefore, this means accessing credible information from participating agents could be an objective for a reputable network. This makes the case for researching into agent behaviour and how it can affect the credibility of messages in online social networks.

2.3.4 Online social swarming

The online community is growing daily with more and more people finding friends or joining in discussions to give their opinions. Even though there have been concerns raised by previous studies evaluating the effects of factors like credibility in obtaining data through the

Internet, it continues to be a popular platform for sharing both personal and public information. Zhang and Leezer (2009) discussed two kinds of network structures featured in social network: *small-world network* and *scale-free network*. Watts and Strogatz (1998) pointed out the small world property of a complex network reiterating that nodes have a small number of connecting edges (relations) with other nodes in the network they belong For example, students in a class operate in this network during their school year and as they move forward to another class or graduate their relationships with members of the class (network) begin to slowly diminish. Barabasi and Albert (1999) discussed the scale-free property of a complex network in their literature and, as a result, many more researchers have continued to research on their findings. A scale-free network is a network where a small group of agents have a high number of connecting edges, meaning they are connected to a large number of agents. The Internet functions as a scale-free network because of its dense population of users.

The process of opinion formation (by agents) in social swarming provides great potential for gathering information in social networks (Kaiser, Krockel and Bodendorf, 2010). Social swarming involves agents having similar content interest and sharing the information gathered with their network. According to Probst *et al* (2010), a social swarm is described as a gather and share technique for social networks. Incorporating *corroboration* is a method by which credible information is attained in this context. Social swarming is a quick and easy technique for different agents to contribute to the database of a social network (Lui *et al*, 2012).

Online social networks (OSNs) are seen as platforms for swarm creation, functioning between social peers and non-social peers, aiding the distribution of information in a network. In the context of social networks, an agent is referred to as a *peer*, also known as a node. Peer-to-Peer (P2P) system is an example of a social peer swarming method for OSN (e.g. a Bit Torrent client) (Zang *et al*, 2012). According to Wang *et al* (2012), Bit Torrent in particular has achieved tremendous success among Internet users. The connection between OSN and P2P data swarming is more like a two-way interaction process, given that in addition to the benefits derived from OSNs on data swarming, OSNs might also be significantly influenced or shaped by P2P data swarming.

A non-social peer is a static single-source model of social swarming, which is a slow mechanism for information gathering within a network. The benefit of social peer when compared to a non-social peer is that the former exhibits faster transmission and dissemination of information. Social learning in complex networks (social networks) is gradually gaining the attention of researchers in social sciences and in ecommerce. This system centres on how social network structure and agent stratagem influence the spread of information and belief formation (Aili *et al*, 2011). *Bayesian* learning, *Non-Bayesian* learning and *Ad hoc* learning are mechanisms for social learning and opinion formation that is related to this research work (Sridhar and Mandyam, 2011).

- Bayesian learning: agents update their current belief based on the actions of other agents or through information obtained by communicating with other agents in their network; it captures the social influence of all connected agents (Aili *et al*, 2011).
- Non-Bayesian learning: focuses on assessing previously held beliefs with social influences to produce a new set of beliefs (Aili et al, 2011).
- Ad hoc learning approach: combines social influence with public beliefs to capture a specific view of connectedness for decisions taken (Sridhar and Mandyam, 2011).

These three mechanisms are the key elements used in the design of this research experiment because they display the characteristics of real world networks. The agents in this research simulation have individual beliefs by way of characteristics, but are able to add on new characteristics if they choose to at random times when the simulation is running. Alternatively they can stick to their beliefs when making decisions at any given time.

2.3.5 Peer-to-peer networks

Peer-to-peer (P2P) networking has generated tremendous interest amongst many Internet users worldwide, especially with computer networking professionals (Mitchell, 2016). The network is fashioned for file sharing between two or more computers connected together without the use of a server computer. According to Nogueira *et al.* (2002), this file sharing power has resulted in Internet connected computers acting simultaneously, as both clients and servers create the P2P architecture. It is used in many organizations around the world and is fast becoming a growing trend for individual users as well. Like many systems P2P network have been defined by many researchers but for the purpose of this research P2P is defined as:

> "*a self-organizing system of equal, autonomous entities (peers), which aims for the shared usage of distributed resources in a networked environment avoiding central services*" (Steinmetz and Wehrle, 2004).

P2P systems have two major architectures: Decentralized and Semi-centralized (Moro *et al*, 2002). They function as agents that act as combined clients and servers, which then introduce the direct exchange of data with other agents as necessary, thereby supporting the P2P application.

- *Decentralized Architecture*: This network lacks a central point of control, making it a traditional P2P system. All the agents (peers) in the network act autonomously, with each being relevant to the network and regarded as being both equal and independent from other agents. This feature gives agents the ability to communicate directly with other agents or indirectly through intermediary agents within their networks. This also means agents can connect and disconnect at will (Moro *et al*, 2002). An example is a user's presence on Facebook where there is no central point or a regulating head; every user is equal, both receiving and contributing to the network.

- *Semi-centralized Architecture*: In this network, unlike the decentralized P2P system, there is a sense of control from one or more agents designed to organize the network, creating easy links to a desired agent if and when needed by other agents (Moro *et al*, 2002). When a node belongs to a group regulated by a head or members of an invitation only network, this is an example of a semi-centralized architecture.

The structural characteristics of peer-to-peer networks have also been included in the research simulation design, using the decentralized architecture approach. This enables nodes to display real world network characters that generate data in that scenario.

2.4 Trust in online social networks

Trust is a factor related to every topic or issue arising within the premise of living things. Humans and animals come across trust daily as an aspect of their being, which contributes to successful interactive phases within their communities.

In the world of the Internet, where humans operate, trust is an utmost factor that should be taken seriously. Mu and Yuan (2010) present trust as a credibility tool used to judge the quality of participants in a service-oriented environment like a social network. The analysis of human behaviour pertaining to trust is relevant for relationship building in social networks. Every contribution made by all participants within the network needs to be reliable to achieve the set goal of the network. According to Golbeck and Parsia (2006):

"*Trust in a person is a commitment to an action based on a belief that the future actions of that person will lead to a good outcome*".

The two key words in this definition are: commitment and belief. Belief alone does not necessarily mean there will be trust. However, trust is created when the belief is the starting place for making a commitment to an action. Trust has also been defined as:

"*a subjective degree of belief about agents or objectives on user's previous experiences and knowledge*" (Kim *et al*, 2008).

Amongst trust characteristics this research identifies the following;

- Direct Trust
- Asymmetric trust
- Transitivity trust
- Personalisation trust
- Reciprocal trust
- Recommender trust

These factors form the foundation from which the Internet operates.

2.4.1 Online trust characteristics

1. **Direct trust**: This is a kind of trust that occurs between two people, such as Ann and Tom below (Mu and Yaun, 2010). They both have a direct relationship with no third parties as an intermediary.

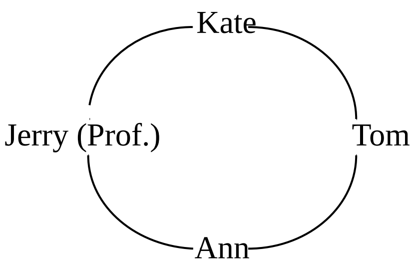

Figure 2.3: Example of the connections between nodes in an online social network.

2. **Asymmetric trust:** is a one way trust system occurring mostly due to ranks, an example being between a professional (Jerry) and the non-professional (Kate), as demonstrated in figure 2.3 above (Golbeck and Parsia, 2006). Kate is of the opinion that because Jerry has attained that position, he should know the answers better. For social networks, however, reputation also comes into play in this context. The idea is that Kate will trust Jerry because she believes he is an expert in her area of interest, having already linked him to a reputable organisation on the web (Mansoury, Shahriari and Shajari, 2012). Asymmetric trust can be viewed in another context; observing the daily life of a mother and child, where a baby trusts their mother without any doubt, and a mother watches her baby closely because she knows that the baby is not to be left alone, especially in the early stages of life.

3. **Transitivity trust**: an illustration of trust is that, if Ann highly trusts Tom, and Tom highly trusts Kate, then Ann may then trust Kate due to Tom’s position in her trust hierarchy. Although this is not a common method within networks, trust could be passed from person to person in some circumstances, especially in the online purchase of goods (Golbeck and Parsia, 2006).

4. **Personalisation trust** is an important characteristic for social networks. Everyone has a personal opinion regarding what they trust or, in some cases, distrust. Therefore two people will have different views on a particular subject issue according to their trust factor.

5. **Reciprocal trust** is trust generated between two parties in exceptional cases. So Ann's trust of Tom will cause Tom to develop some level of trust towards Ann (Wang, Wang and Lui, 2002).

6. **Recommender trust** encompasses all the aforementioned trust characteristics. Recommendation within social networks works solely on a user's reputation, it denotes the objective opinions of the user's expertise from online community members (Kim, *et al* 2008).

2.4.2 Evaluating trust in online social networks.

Reputation mechanism and Incentive-Compatible Escrow Mechanisms (Witkowski, 2011) are constantly being developed to improve the social networks. Many e-commerce site rely on feedback to market trust in social media. However, this feedback can be questionable as it cannot always be relied on. According to Mansoury *et al* (2012), the assumption of truthful buyer feedback is unrealistic because people's opinion vary and the fact that a buyer is not obliged to the seller especially when there are no incentives to persuade the buyer to take time to give feedback, dishonest reports cannot be avoided. Escrow mechanism entails having an intermediary between a seller and a buyer, in this instance the buyer does business with the third party company and, therefore, gives their feedback directly to the third party as well. This method eliminates the problem of *whitewashing*; the ability to create new profiles (Cho *et al*, 2012). This is the process that goes on between buyers and sellers using Ebay for their many transactions daily. Ebay is the third party company providing the platform for buying and selling of goods and services.

Reputation mechanism can generate distrust within social networks because most of the networks have large membership strength and it is unlikely that people will know each other (Victor *et al*. 2011). Therefore, for example in figure 2.3, if Ann needs to clarify a fact in a book written by Jerry and she confides in Tom, Tom may not know much of that field and so will then seek the help of Kate. Within this network the level of trust can be questioned and Jerry's reputation could either be boosted or dented. This could bring about a case of collusion between friends and the creation of false identities undermining the reliability of node reputation (Hogg, 2009).

2.4.3 Reputation mechanisms

Social network messages use reputation mechanisms to decide the credibility status of messages generated in a network. It functions in two ways, according to Wang and Chui (2008), one in automatically identifying reputation based on position in the network and the second: as a filtering tool to guide a user's rating of information. In e-commerce, online trading usually relies on trust (Hogg, 2009), also according to Klein (1997), repeated transactions between sellers and buyers is encouraged by reputations harnessing truthful behaviour reviews and customer ratings, as is the case in many online market places e.g. eBay. Companies request customer reviews to assist other customers to make good decisions, but these mechanisms can and have been manipulated by collusion among friends creating false identities or multiple identities (Resnick *et al*, 2000). Friends can give each other high ratings despite poor performance, which will give a new customer incorrect information that may influence their choices.

The theory of 'Six Degrees of Separation' is an experiment carried out by Yale university professor *Stanley Milgram* in the 1960s to test what he called the 'small-world problem'.

This test illustrated that two strangers were connected by six other individuals between them, beginning with one of their friends.

> "*It points out that any two strangers would be able to recognize each other through no more than six people*" (Xiao-hua, 2010).

The result of this theory is has contributed greatly to social media, showing the close connections that exist between individuals and has influenced researchers study of social networks, with particular reference to how user connectivity continues to expand. In recent times, however, with digitalization around the world the continual growth of the Web, modern society's connections and capabilities have changed and continue to change each passing day.

In other to achieve any kind of credibility for information processed through the Six Degrees of Separation theory, information is carefully analysed by individual nodes before a decision is made because messages are easily modified. With online social networks this could be challenging, because of the number of messages passing through the network leaving less time for message validation, hence the reason for this research; finding possible ways agents can identify credible information.

2.5 Credibility

The growing popularity of online social networks has made information sharing and innovation not such a difficult task now, compared to what was obtainable in the past. Over the years, social networks have become a participatory culture spanning from creativity, contribution to collaboration. It has broadened its network from merely being an information provider to an interactive entity (Chandrashekar and Hockema, 2009). According to Canini *et al*. (2011), social networks encourage their members to simultaneously become both consumers and contributors of content in a network. When this happens, the roles of the members extend from just a few contributors to a more diverse group of individuals. This process increases data sharing. However, this then creates a platform where anyone can produce content for a worldwide audience, irrespective of credibility status (Chandrashekar and Hockema, 2009).

Credibility in computing was first discussed early in 1999 by Tseng and Fogg in an ACM journal. Here they asked questions relating to the use of computers;

> *"What is credibility? What makes computers credible? And what can we, as computer professionals, do to enhance the credibility of the products we design, build and promote?"*

They defined credibility as believability, stating that those who are credible are believable people, which could mean that credible information is information that is believable (Tseng and Fogg, 1999). From their research findings, they then proposed four types of credibility: Presumed, Reputed, Surface and Experienced. They also highlighted that psychologists at the time had only outlined factors that contributed to computer credibility, making it necessary to have these four categories aimed at providing new ways of evaluating (thinking about) computer credibility. As the use of computers has developed over the years since 1999, types of credibility have evolved from just a standalone tool to a communication tool into messages or information distributed throughout social networks in every form available, including print media, videos, instant messaging, blogs and online social networks.

In 2012, Lewandowski defined credibility as:

> "*Peoples' assessment of whether information is trustworthy based on their own expertise and knowledge*".

This was created as a result of information consumption, which placed it in the hands of the consumer. Credibility is a crucial factor for the growth of the e-commerce market hence it is necessary that it attains a reliable network environment. As mentioned earlier, according to Tseng and Fogg (1999) '*credibility is believability*' meaning information is termed credible because a user believes in it. If there are no people to believe in a message, then that message fades away as there are no people to share it. Believable information is obtained in three ways, as stated by Liu *et al*. (2010):

- *Homophily*: likeminded belief.
- *Test-and-validate*: recipients of information research the truthfulness of the information retrieved.
- *Corroboration*: a continued reporting of information by several sources.

In the case of social networks, because users tend to have *reciprocal trust* for one another there is the assumption that they should find credible information regarding their area of interest therefore confirming the definition laid out by Tseng and Fogg (2009).

There are some concerns regarding the credibility of socially networked information. The information Naiveté, as proposed by Brody (2008) is briefly examined below:

> '*Knowing of* and *knowing about*'.

Many users are grouped into the first category:

- *Knowing of* is where users have information on certain topics but are not totally knowledgeable about them but place the information out there regardless, for other users to decide. This kind of approach thrives within social media because users, desperate for information, seem to encourage what this research terms as ‘get-quick information’.

- *Knowing about,* on the other hand, encourages publishing and stands for detailed knowledge of a given area of interest, placing credibility at the top of its existence (Brody, 2008).

There is also concern regarding influenced opinion for the creation of information on a particular subject for social networks. People’s views can influence others when they make their contributions in a group discussion (Lui and Shakkottai, 2010). A good example of this is when people choose their mobile phone operators; they look for the networks that are used by friends and family to get incentives provided by that operator such as, for example free calls to the same network.

2.5.1 Types of credibility

There are four types of credibility as outlined by Tseng and Fogg (1999) for computer credibility;

- Presumed Credibility
- Reputed Credibility
- Surface Credibility
- Experienced Credibility

This research has focused its discussion to messages generated in online social networks. What each means is discussed below:

1. **Presumed credibility**: This portrays how people take in information created by an individual, out of assumption such as, for example, when an article is published on a BBC web page, it comes with the assumption that the editor is a renowned journalist in a reputable organisation and, as such, is trusted to report only credible information (Tseng and Fogg, 2009). Referring to the example given by Tseng and Fogg (1999), sales people are said to lack credibility, hence the assumption that car sales people are not to be trusted. More specifically, when an individual wants to buy a car they seek advice from friends about their choice of car or take a friend to the dealers to help make a good choice presuming that the friend in this case is trustworthy.
2. **Reputed credibility**: This is a case where third parties have tagged a web page with credible status, which in turn prompts a user to believe in them (Tseng and Fogg, 2009). The reputation of a person as a professor, for example, makes people believe they are a credible source of information, or where the professor has been given a prestigious award based on research they have carried out. To further understand this concept, here is another example: when shopping for a product online that is branded, credibility is attached to it based on the company name. Most people would rather buy a branded product than one that is unbranded even though it may never be known whether the product lives up to the brand name, unless it is tested with one that is unbranded. However, a test is not carried out because the company whose brand the product bears is trusted, even though the manufacturer is not known. This can be likened to *recommender trust*, holding on to a third party's report to pass judgement on a subject.
3. **Surface credibility**: In this type of credibility, the phrase 'don't judge a book by its cover' is referred to, but rewritten as 'judging a book by its cover' (Tseng and Fogg, 2009). Users of the web assume because a web page is well designed, it is

a credible source of information. In everyday life, this can be illustrated as giving people credit for their looks or language of communication. Many holidaymakers have experienced this phenomenon when they have seen pictures of hotel rooms online but, on arrival, all turns out to be a lie, even though in most cases very good reviews have been left in the review sections online.

4. **Experienced credibility**: This kind of credibility believes in something based on first-hand past experience (Lewandowski, 2012). For example, a stockbroker's previous success with their client builds up credibility for their business in the client market. When information retrieved through a search engine is proven true over time, credibility status is built up for future users.

These credibility types are illustrated in the form of a pyramid in figure 2.4. Each credibility type has been placed at different levels of the pyramid according to their definition:

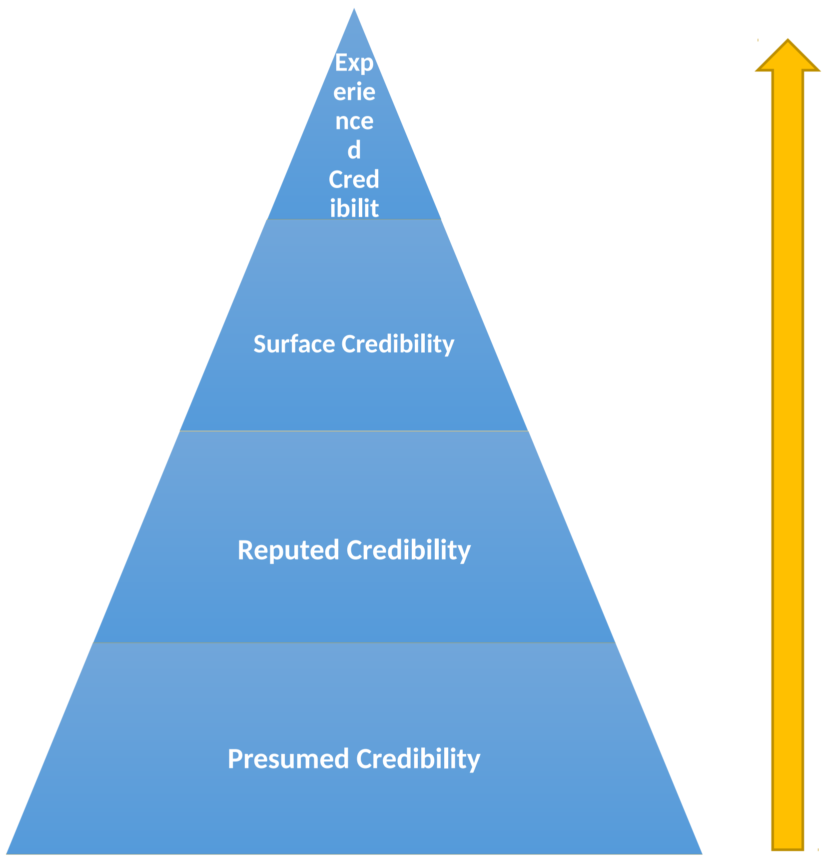

Figure 2.4: Credibility pyramid showing the different stages of nodes in a network based on messages transmitted over time.

2.5.2 Credibility Factor of online social networks

Social networks gain credibility through recurring searches by different users at different virtual locations geographically. This research proposes that the amount of time spent by a user on a webpage is registered in an invisible personal credibility log. Studies have shown that the more users use a network, the more they become dependent on said network, which therefore means they give credibility to the resources generated (Bian, 2012). This also means that such users are more likely to suggest this network to their friends. Another credibility factor in social networks is the mechanism of reposting or as *Twitter* puts it, 're-tweeting'. Liu *et al*. (2012) suggest social swarming, in which case users armed with smartphones are directed by a central director to report on events in the physical world. This technique generates a certain state of credibility for the online network as it functions with *corroboration*.

One social factor described by Bian (2012) estimates that people with similar faiths, social lifestyles or economic capabilities tend to have a common likely behaviour as to how they think, which aligns with the term 'birds of a feather flock together'. Users, who find themselves in this group, build credibility based on their flock influence or preference regarding the data they rely on, as the case may be.

2.6 Collaboration in online social networks

A social network structure is designed so that it functions properly only with agent interaction and communication. Apart from known social networks such as Facebook and Twitter where message generation and sharing is facilitated, there is another kind of network called collaborative network. According to Miao *et al*. (2012), agents in these networks cooperate to complete a given task. For example, in online transactions, agents are faced with the task of resolving customer queries or fixing website failures. In these scenarios, all efforts are geared towards improving customer service.

Sometimes, other agents in the network could influence an agent's behaviour. As a result, a lot of the behaviours are not what most agents would likely showcase if they were in an isolated environment. In other words the behaviours of other agents is most likely to be the reason for many decisions taken by individual agents across a network. To form a network, individuals interact with each other to form strong bonds with the intention of benefitting mutually. Within many networks, several interactions are typically based on service transactions were agents desire to meet the demands of other agents (Zhang and Lui, 2006).

The interacting social nature of software agents presents a challenge in multi-agent systems, such as the process involved in securing message distribution in a network with emerging computer-based applications for online businesses.

Collaboration in networks is based on *Trust* and *Credibility*. Trust can have a negative impact on collaborative networks. This is because when one agent defaults or is manipulated, it may be challenging to detect the one agent responsible. In this research area, these agents can disguise themselves in a manner aimed at deceiving other agents. Not all agents, however, have malicious intent as some contribute positively to completing the task in the network.

Reputation-based trust and referrals are still major concerns relating to this area of study. In the research findings of Porter and Sen (2007), exploration strategy for rewiring provides more benefits over random selection; their findings go a step further by smoothing out the process, thereby decreasing the estimated number that may have been required previously.

The act of rewiring, according to Barton and Allan (2007), is one of the activities agents undergo in a network to achieve their daily task and includes:

> *“joining an existing team, initiating a new team, waiting, or rewiring the connections”*.

This phenomenon allows agents to break free from any relationship that they feel is no longer beneficial, thereby substituting such connections with new ones (Villatoro *et al*. 2011).

2.6.1 Collaborative networks

A collaborative network is a kind of online social network that is formed specifically by members to collectively achieve a task and could be termed ‘private’. Such networks are driven by the information circulated among its members for the soul purpose of completing the task assigned to the network. Unlike in public social networks, where information can be generated by any agent and which spreads rapidly into the network through other agents (Bentahar *et al*. 2007), information flow in this network is powered by certain tasks and generated by one source agent, which then makes its way into the network until it locates an agent that possesses the expertise to carry out the tasks

A major step in achieving a collaborative network is first to search for trusted agents. This process is influenced by the many characteristics of agents, one of which is autonomy: self-controlled agent communication in an open environment with free access to a wide range of resources from which to choose. Since agents sometimes rely on the reputation knowledge of other agents for interaction, many malicious agents can pretend, for the sole purpose of gaining trust, to be chosen. After this they begin their activities penetrating into the network having gained some level of credibility and, in order for information to be shared and utilized, interaction must take place, thus allowing for mutual access of resources (Bentahar et al. 2007).

Rewiring is a major instrument for the propagation of information within any given social network. Therefore, in the case of collaborative networks, it is not different and is a social instrument:

> "*that facilitates the emergence of norms from repeated interactions between members of a society*" (Villatoro *et al*. 2011).

2.6.2 Rewiring

Rewiring can be achieved by one of these procedures:

1. Random Rewiring: Selecting agents randomly from a vast population of agents in a public network.
2. Neighbour Advice: This is based on reputation and referral.
3. Global Advice: Choosing an agent according to certain criteria or strategies that have been used in the past.

The concept of rewiring is important for online social networks at all times because it forms the basis for continuity in network interactions and growth. However, in the process of connecting with new agents, agents in a collaborative network can be tricked into welcoming pretentious agents into their network. Rewiring is an uncertain process network agents have to carry at different times. Therefore, when choosing trusted partners, utmost care must be taken. A task distributed in a collaborative network can either be altered maliciously by some agents or shared outside the network without permission, depending on how important the task is. There are many cases of companies that have fallen into the hands of such agents in the real world resembling a case of corporate espionage. In situations where the network fails to notice the presence of such activities, it is likely that it will continue undetected.

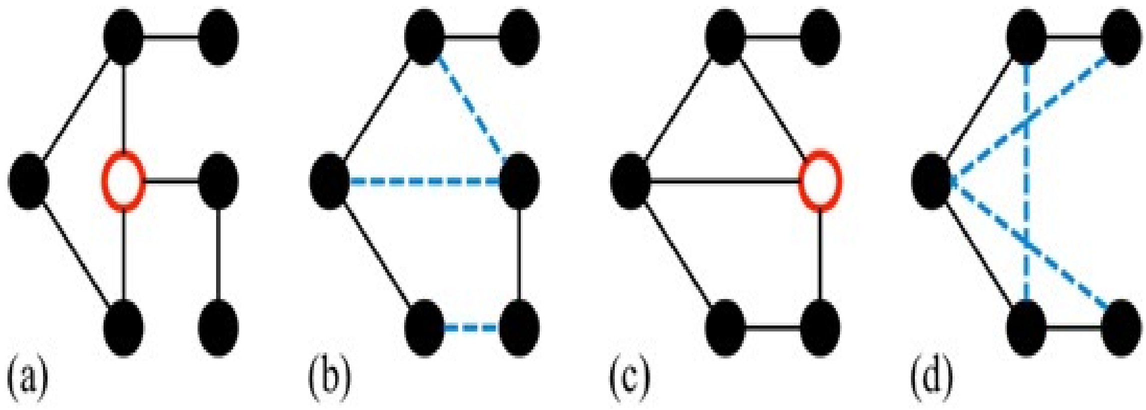

Figure 2.5: Network adaptation with random rewiring: (a) shows the initial network with the targeted node outlined, (b) shows the adapted network following the node removal with rewired links shown as dashed blue lines, (c) shows the next targeted node and (d) shows the subsequent adaptation (Tran *et al*, 2016).

Rewiring is used to disconnect from agents deemed no longer of interest as a result of either dormancy or interest collision. This is the first step in the process of rewiring. As new acquaintances are made a friend request and acceptance takes place to complete the process. Within a large network like Facebook, there are groups that have been created which may also have subgroups, essentially created by people with a more streamlined interest, such as, for example, the Facebook page owned by the University of Derby and that owned by the Student Union at the University of Derby.

It is likely that many of the members of the university group may not be in the student union group and vice versa. In this case, the student union group is the collaborative network. Rewiring can take place in this network as often as necessary, but may be more extensive when new students enrol into the university. Agents in the student union group can connect with the university group to connect with other agents that may have joined the network more recently in each academic year.

Rewiring frequency can be ‘defined’ as ‘dynamic’ in nature based on an agent’s or a group of agents’ interest. This could either increase or decrease rewiring accordingly. In this state, there will be fewer agents in a network with a dynamic frequency as compared to a defined network, with more in the public network. In the illustration above, Social Interaction represents closest interest with the agent; there is more communication happening in this group (the student union group). If the group section is a network connection of ‘not so close interest’ but ‘defined’ (the university group), then the world is the public Facebook network.

Social activities in a network can influence the frequency with which agents make new connections. There can either be an increase or a decrease in the number of new connections. This is determined by the volume of activities carried out in the network, meaning the faster each task is completed, the quicker the new tasks are introduced, making room for possible rewiring. This is to connect with agents that have the desired skills beneficial to the most current task introduced into the network. Understanding the fundamentals of a collaborative network will determine how information passed from agent to agent can be considered credible. This is important not only to the agents with shared interest, but also to those in the public network.

2.7 Opinion formation and the challenges affecting messages

Instant messages are shared daily, by the second, using online social networks. This is mostly carried out between strangers and acquaintances in the real world, but with friends in the invisible world of the Internet. In a micro-blog such as a tweeter, re-tweeting is a feature available to users for instant sharing of an interesting message posted by other users for the benefit of their relationship circle or group. Malicious nodes can share messages through this feature in a network and it can go undetected, causing other nodes to make decisions based on such messages.

The discussion of credibility and trust leads to suggesting how one can sometimes lead to the other. It is generally expected that some nodes will first trust a message before recommending it to friends within their sphere of contact, therefore passing on credible information. Opinion formation is a great feature associated with how a user makes a decision to trust a message before proceeding to sharing it with other users. Common interest brings people together, so it is in the online community where agents build relationships with many other agents enabling new forms of social organization. According to Kolbitsch and Maurer (2006), online networking or interaction leads to changes in opinion formation when information is exchanged. The use of social network analysis to examine relationships amongst interacting agents is usually applied to social networks online for the analysis of opinion exchange (Kaiser, Kröckel and Bodendorf, 2011).

Credibility refers to an individual degree of convincing of a certain thing or fact in a system. In the online community, especially in social networks where information is rapidly and continually shared, there are some factors that impact online credibility (Bian, 2012). Those related to this research work, and based on opinion formation, are discussed below:

1. **Social Factors**: This refers to the social class of the agents in the network. Most of the time, users in a network are, for example, of the same faith and economic capability. With this in mind, it is expected that their mode of thinking and how they respond to information while giving their opinion can be influenced by their beliefs. In this network structure, because of the class factor, credibility to judge or contribute information is widely influenced by one member to another. Over time

members of this network form reference groups (people met outside the internet or often kept in touch with) and these subgroups in the network form their own opinions based on their relationship.

2. **Information Types**: Network information is classified into four types: news information, business information, entertainment information and reference information (Bian, 2012). News information covers mostly political events, business news focuses on purchased products related to a company, entertainment news refers to recreation information providing entertainment programs e.g. movies, music or games while reference information relates to information gathered from the internet by users and referred to in real world activities. Bian's report contains an experiment, that was conducted to illustrate the credibility status in these aforementioned categories of information, the results of which are shown in table 2.1 below:

Type Frequency	News Information	Business Information	Entertainment Information	Reference Information
Utterly Incredible	17.30%	25.70%	21.40%	10.70%
Uncertain	54.40%	55.90%	56.20%	55.20%
Fully Credible	28.30%	18.40%	22.40%	34.10%

Table 2.1: Experimental evaluation data (Bian, 2012).

Amongst the four types from the above table, news information and reference information both have a higher credibility percentage compared to that of business information and entertainment information. This could be because entertainment information and business information are both geared towards profit making, therefore giving room to personal opinion formation and meaning that, whatever the giver of the information wishes, is what he/she would post online to generate profit.

3. **Information Input/Involvement Degree**: The majority of online network communities require a user to give personal information before they are allowed to access information available in the network. Because the issue of safeguarding one's identity is now commonplace, many users either create false identities or keep personal information private. When such identities are created, because many user accounts or profiles are not credible, what then can be expected regarding the credibility of message that are posted by such users?

2.8 Summary of this chapter

This chapter has discussed the many areas relating to the theoretical view of the credibility of online social networks. What agents are and how that they have the ability to function autonomously in social networks was discussed in detail. It is important to note that the abstract nature of agents is the key rationale for using them. Therefore, this research has used agents to help represent the complexity of online social networks by adopting the agent-based modelling approach. Emphasis was made on understanding what a network is with figures demonstrating what happens in an agent network. Trust factor for agents (direct, reciprocal and recommended trust) was also discussed. Agent credibility and the four types of credibility were also evaluated, referring to their use in other studies and how this highlighted these four into a credibility hierarchy.

This research focuses on how agents behave towards messages in their networks and what category of credibility they attach to the messages based on their characteristics. This assists in placing messages on the hierarchy, whether they fall under Presumed, Reputed, Surface or Experienced (PRSE) credibility. The procedures and topics discussed in this chapter are relevant in the design and implementation of the research experiment.

Chapter 3

Methodology

3.0 Research Method

This Chapter sets out the method used in the research. The researcher also discusses the following areas:

1. Discussion of the methods of research to be used; quantitative or a qualitative approach.
2. The specifications for the experiment and expectations in implementing these specifications.
3. Rationale for and evaluation of the design of the experiment.

3.1 Methodology

A research methodology is a framework within which a research study is conducted. It determines the approach to be used in providing answers to a question; this approach can be put into practice for a research process (Remenyi *et al*, 1998).

The method to be used in a research study must be outlined clearly in order for the results of the research to be given credibility or be convincing. The purpose of the research and the research question will most likely determine the choice of method. Qualitative and quantitative methods are the two major types of methodology used today. They operate through different approaches, depending on the underlying goal the researcher intends to achieve. According to Cassell and Symon (1994), qualitative and quantitative research is each appropriate for different types of research problems, which means the research question determines which route to adopt in order to achieve the anticipated results.

3.1.1 Qualitative research method

Qualitative research provides an in-depth description and understanding into human experience. Its main purpose is to describe and understand the human phenomena, interaction or discourse (Lichtman, 2010). Qualitative research has generally (though not exclusively) been associated with some sets of beliefs, including:

- *Perception relating not only to the senses but also to human interpretations of what the sense imply.*
- *Knowledge of the world is based on understanding that arises from thinking about what happens, not just simply from having had particular experiences.*
- *Knowing and knowledge transcend basic empirical enquiry.*

- *Distinction that exists between 'scientific reason' (based strictly on casual determinism) and 'practical reason' (based on moral freedom and decision-making which involve less certainty).*

  (Ritchie *et al*, 2003)

Those applying qualitative research tend to place emphasis and value on the human interpretative aspects of knowing about the social world and the significance of the investigator's own interpretations and understanding of the phenomenon being studied (Ritchie et al, 2003). The phenomenon deals with human experiences, interactions with other humans and the processes of the communication that takes place. Generally speaking, these three features can be intertwined when researching on human behaviour (Lichtman, 2010).

Although there is no clear definition of qualitative methodology, Remenyi *et al* (1998) have defined it as:

> "*the methodology based on evidence that is not easily reduced to numbers*".

As described by Lichtman (2010), a qualitative methodology:

> "*covers an array of interpretive techniques which seeks to describe, decode, translate and come to terms with the meaning not the frequency, of certain more or less naturally occurring phenomena in the social world*".

The researcher collects, organises and interprets data, with his or her eyes and ears as filters. These data are subject to a host of interpretations, even though there is no single interpretation that is better than another (Lichtman, 2006). The underlying objective of a qualitative researcher is to describe and understand human behaviour beginning with assuming or knowing something.

3.1.2 Quantitative research method

Quantitative methodology identifies specific variables or looks for relationships amongst and between different variables. The aim in this circumstance is to look for how one or several variables are related together using data sets. It deals with hypothesis and how to test said hypothesis with numeric data in measurable quantities (Lichtman, 2010). This approach is experimental and focuses on the outcome rather than the process used to get the result. In most cases, it is statistically justified, which makes it rigidly defined and less flexible. It usually happens in an experimental setting; as such when a single correct answer is required (Cassell and Symon, 2004).

Some basic comparisons between Qualitative and Quantitative Methodology:

Qualitative Research	**Quantitative Research**
Is concerned with behaviour and situation	Concerned with cause and effect
Is focused on Interpretation	Focused on quantification
Gives room for flexibility and less rigidity	Less flexibility and rigidly defined
Longitudinal research design	Cross sectional research design
Emphasis on the richness of qualitative data	Emphasis on statistical data
Emphasis on subjectivity	Emphasis on objectivity
Numbers may not be of importance but might be involved in different instances	Involve numbers most of the time
Treat those studied as participants and informants	Treats those being studied as anonymous objects to be measured
Takes place in naturalistic settings	Takes place in experimental settings
Uses inductive approach	Uses deductive approach
Progresses from specific to general	Progresses from general
Works with observations	Involves sampling and surveys
Understands and interprets the meaning of human interaction	Statistical test to decide whether or not to reject a null hypothesis
Relies on interpreting and understanding behaviour	Relies on hypothesis testing and analysis
Is geared towards process rather than outcome	Is geared towards outcome rather than process
Is concerned with emergent themes	Is driven by specific hypothesis

Table 3.1: Comparisons between Qualitative and Quantitative Methods in Research (Lichtman, 2010; Cassell and Symon, 2004).

3.1.3 Qualitative and quantitative mixed method

This research mostly uses the qualitative method because the research question deals with the subject area of understanding and interpreting agent behaviour within a given network, even though it was anticipated that the result generated by the simulation would have some quantitative data because nodes were represented in numbers in the network (agents are numbered from 0 to as many that may be required). This means the resulting data recorded after a simulation run would also have some numbers. It is not uncommon to use mixed methods in a research project, and this just shows the diversity of the topic in question. The appropriate method was used at the required stages in this research.

Qualitative methodology enabled the researcher to generate randomly assigned data based on agent behaviour, evaluate and understand the characteristics that individual agents displayed and the reaction of other agents to single agent behaviour. A simulation was designed to replicate the natural environment of agents in online social networks to enable flexibility and understanding of the process involved.

The researcher analysed data generated from the simulation by highlighting the behaviour of nodes in the results using the credibility hierarchy defined in the literature review.

3.2 Research Design and Approach

A network simulator is a practical method for producing data from imaginary agents and conducting analyses on the data generated, thus imitating real life scenarios. It cuts down the amount of money and time that would have been invested in an expensive process or software component (Brakmo and Peterson, 1996). It requires the researcher to feed the simulator with the different characteristics needed to produce results that are then subjected to comparisons. These simulators create a virtual environment of nodes and edges in a network. A virtual environment gives room for manipulation of characteristics, especially when dealing with sizeable networks. The purpose of using a network simulator is to generate a range of reports to be used in statistical analysis aiding the understanding of agent behaviour and communication patterns.

3.2.1 Agent-based modelling simulation

This is a simulation approach spanning over 15 years (Knoke and Kuklinski, 1982) used for dealing with complex studies of individual agent behaviours and interactions. Agent-based modelling and simulation (ABMS) has been applied in many research areas, ranging from biological systems, agriculture, epidemiology, market analysis, social networks, crime analysis and evacuation. The main objective of ABMS is mostly abstract, which means it is almost impossible to understand detailed description of an actual human behaviour. The simulation analyses the simultaneous actions and interaction of multiple agents, with the goal of reproducing and predicting the outcome of complex phenomena. To create and run a model, it is necessary to use software, enabled by a programming language or other implementation techniques that offer this capability.

Autonomy gives room for unsupervised interaction, meaning that an agent can exchange information with another agent independently during this process to achieve its goal within an environment. Agents are seen to be self-contained and having certain characteristics or attributes, certain behaviours, and autonomic decision-making abilities. They also share moderate features that set out boundaries, making it possible for predicting whether something is part of an agent or not, or whether it could be a feature shared by other agents.

Agent based modelling also allows the introduction of different parameters into the simulation, at different times, to produce different results.

3.2.2 Preliminary approach

In order to determine the credibility of the messages generated via social networks, the model presented in this research aimed to use ABM to simulate the spread of messages in a social network. There were a number of steps that were taken. The first identified the structure of the created network. The second generated different attributes (e.g. age bracket or profession) for the agents and creating rules that would alter the behaviour of agents in the network. Random messages were introduced into the network intermittently, in order to aid the analysis of the mechanisms that defined the message movement. It was intended that the ODD protocol would help in formulating the concept for the simulation.

3.3 Overview design and details protocol

The Overview, Design concepts and Details (ODD) protocol is a generic format and standard structure by which all Agent Based Modelling Systems can be documented (Grimm *et al*, 2006). Published in 2006, it was created to standardize the published description of individual-based and agent-based models (ABMs) and its main purpose is to make model description understandable. The figure below illustrates the main elements of the protocol.

Overview	Purpose
	State variables and scales
	Process overview and scheduling
Design concepts	Design concepts
Details	Initialization
	Input
	Submodels

Table 3.2: The seven elements of the ODD protocol, which can be grouped into the three blocks: Overview, Design concepts, and Details (Grimm *et al,* 2006).

In 2010, Grimm *et al* published an update for this protocol. This was done, not to overhaul the main elements but to make the elements more explanatory because they were being misinterpreted. Experience was cited as one of the reasons an update was required, as the benefits of ODD in research was expanding within the science community. In the years that ODD has existed, it has been used in over 50 publications and continues to grow amongst researchers.

In table 3.2, an updated and listed copy of the ODD protocol is illustrated. The names of two elements were modified (elements 2 and 6), one design concept was renamed (from Fitness to Objectives), and two design concepts were added (Basic principles and Learning). Numbering the seven elements when using the protocol is optional. The elements can be grouped into three categories (Overview, Design concepts, Details; hence ODD), but these categories are not meant to be included when the ODD protocol is being used (Grimm *et al*, 2010).

	Elements of the original ODD Protocol (Grimm et al., 2006)	Elements of the updated Protocol
	1. Purpose	**1. Purpose**
Overview	**2. State variables and Scales**	**2. Entities, State variables and Scales**
	3. Process overview and scheduling	**3. Process overview and scheduling**
	4. Design concepts	**4. Design concepts**
		- Basic principles
	• Emergence	**- Emergence**
	• Adaptation	**- Adaptation**
	• Fitness	**- Objectives**
Design concepts		**- Learning**

	• **Prediction** • **Sensing** • **Interaction** • **Stochasticity** • **Collectives** • **Observation**	**- Prediction** **- Sensing** **- Interaction** **- Stochasticity** **- Collectives** **- Observation**
Details	**5. Initialisation** **6. Input** **7. Sub models**	**5. Initialisation** **6. Input data** **7. Sub models**

Table 3.3: The seven elements of the original and updated ODD protocol (Grimm *et al*, 2010)

3.3.1 Experiment design with the ODD protocol

1. **Purpose**

   This research model was developed to demonstrate the credibility of messages in a network: Under what conditions do agent behaviour and interaction patterns change with the introduction of messages into a network? In what way does the variability of agents affect the spread of messages endorsing credible messages?

2. **Entities, State Variables and Scales**

   The model has two entities: the agents and the messages they disperse. The agents are characterised by many different variables such as sex, location, age, religion, social status, marital status, educational qualification and job titles, amongst others. Each tick represents the change in behaviour towards the message introduced as a result of random interaction. The simulation lasts for 1000 ticks and result forecasted should be the number of accepted messages per individual agent (have-seen agree) measured against the rejected messages per individual agent (have-seen, don’t agree).

3. **Process Overview and Scheduling**

There is a list of processes involved in the model, including: random interaction patterns amongst agents and introduction of messages at different time intervals. At every tick, each agent undergoes various stages that follow a particular order, as outlined in the design concept.

4. **Design Concepts**

- **Basic Principle:** The basic principle addressed by this model is the concept of credible information within online social networks. This concept focuses on message distribution throughout a network after being introduced by an individual agent. The network itself is made up of nodes numbering up to 5,000. Each node represents an Internet user and the behaviour of individual agents towards the messages introduced into the network is monitored. The adaptive behaviour is modelled virtually to suit the interaction of real users in online social networks; randomly connecting to other agents. This behaviour is based on the understanding that real users interact randomly in a given time frame.
- **Learning:** Learning occurs as the agents receive messages; they accept, reject or pass on messages as the simulation moves on.
- **Prediction**: There is no prediction in the model at this point.
- **Sensing:** Sensing is a fundamental process in the model. Agents behave differently according to the network group they find themselves connected to at certain times. In religious groups, for example, certain beliefs alter the acceptance of messages even though the message may be credible; hence there is no tendency to pass on such messages. The model is perceived as an

emerging structure because the resulting data is analysed to develop a framework for determining credible information.

- **Interaction and Stochasticity**: Both of these occur with every tick during the simulation process. This is carried out directly between agents. It is also structured randomly as there are no predetermined numbering configurations or patterns for agents to follow while interaction is going on. Non-directional lines connecting agents together represent interactions and as this continues, their communication is represented by a change in colour, so as to signify their status.
- **Collectives**: Agents form groups or belong to one of many online networks and colours, imposed within the model during the initialization state, represent these social groups. During the simulation process it is anticipated that the collectives emerge, over time, through rewiring.
- **Observation**: Observation is data collected through graphical display. This observation measures number of nodes over time. The number of nodes include agents that have-seen, not-seen, don't agree and shared.

5. Details

- **Initialization**: The message dissemination process is initialized when the model starts. Firstly we will introduce some messages into the model to see the reactions of nodes and continue to introduce new messages randomly over time. The state of the model is set to configure the message spread to agents at 50%, which increases over time. Initialization value is always the same at the start of each simulation process because the values are based on characterised data.

- **Input Data**: The model makes use of external data. This comes in the form of agent characteristics varying at the start of each simulation. These variables change over time and are determined by agent behaviour.

3.4 Simulation Design

The model in this research shows the spread of information in a network case study, based on generation of information by certain agents within the network. In the created network, there are 1000 nodes with varied characteristics for agree or don't agree.

There are instances of some agents in the network not having any contact with the generated message, which is what happens in many social networks. Using Twitter, for example, out of many of its social networking sites, most messages generated end up within the social contact of an agent, sometimes depending on its geographical location.

The network begins with nodes that have not seen a message, and with selected random nodes that are generators of this message.

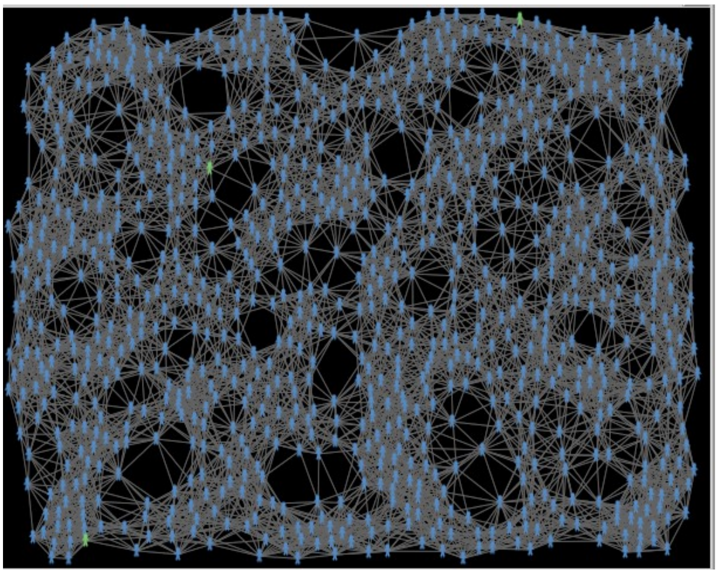

Figure 3.1: Screen Shot A - Agents in a virtual network randomly connected with edges in a setup stage of the simulation process.

The agents have been grouped into six categories and are temporarily represented with different colours for easy recognition. The characters include:

1. Generators (green): These agents have been tasked with the goal to generate the information that will be distributed in the network. The number of generators can be increased or reduced using a slider to observe how they affect the network.
2. Not-Seen (Blue): This is the initial state of every agent at setup, which is the case in social networks when there is no new information. This state continues for some agents who may not come across a message being distributed.
3. Have-Seen (Pink): The agents who are pink are those that have seen the message and decide autonomously whether to share or not to share.
4. Have-Seen-don't-Agree (Red): These agents have seen the message but have decided they do not agree or trust the message and will not pass it on.
5. Have-Seen-Agree (white): The have-seen-agree are agents that have received the message and have agreed to it, but do not trust it enough to share.
6. Have-Seen-Shared (Yellow): Agents in this category have received the message, agree and go ahead to share with other agents.

3.4.1 Evaluating the model

Having outlined the six groups in the network it is important to note that each characteristic is a stage in the network. As a result, one stage can lead to another stage or, simply put, a change in state creates the next category. For example; over time agents that are not-seen become Have-Seen, and thereafter can become Have-Seen-Don't-Agree or Have-Seen-Agree depending on individual decisions. The same can be said for Have-Seen-Shared. This begins from not-seen, to Have-Seen then Have-Seen-Agree before becoming Have-Seen-Shared.

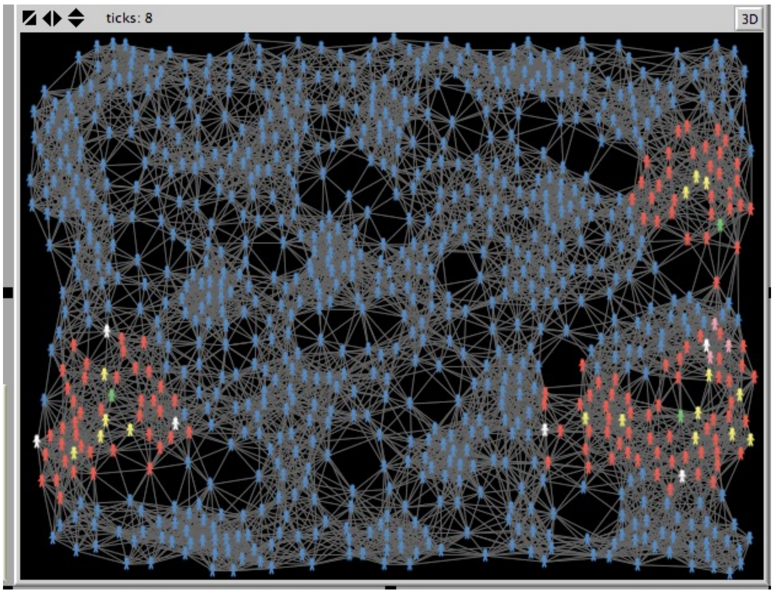

Figure 3.2: Screen shot B showing node changes after simulation run at tick 8.

In screen shot B, the agents display their different stages at tick 8, from blue to pink, then becoming either red or white, and then becoming yellow. The pink agents go from pink to white to yellow (Have-Seen-Agree to Have-Seen-Shared). As the ticking continues, new changes continue to surface; this is illustrated in figure 3.3 below:

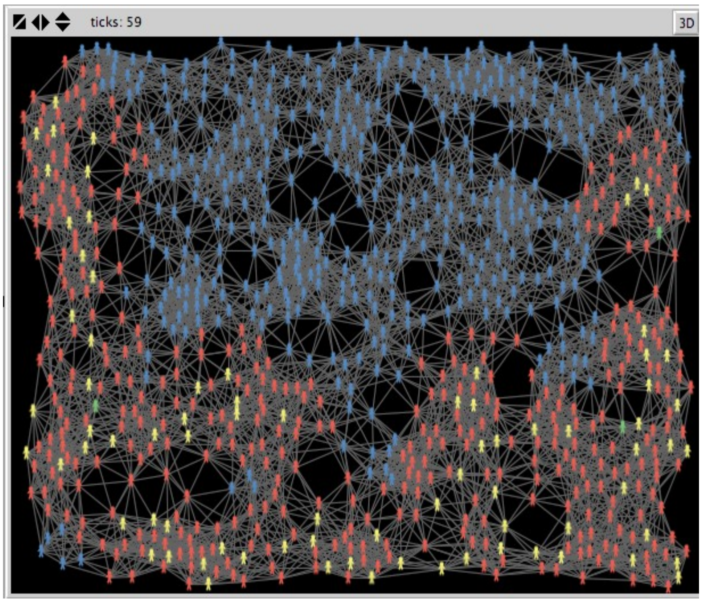

Figure 3.3: Simulation Screen shot C shows node changes at the end of a simulation run stopping at tick 59, seen in the top left corner in the interface.

At tick 59 in figure 3.3 many of the agents are now decided, leaving only 3 categories:

- The Have-Seen-Don't-Agree (Red)
- The Have-Seen-Shared (Yellow)
- The Not-Seen (Blue)

Some variables in sliders have been added to the model to examine how each one affects the behaviour of the agents. These are subject to change as the simulation restarts. Variables created include:

- Number of nodes
- Average node size
- Number of generators
- Likely to agree
- Likely to share

With these variables, questions can be generated and answered accordingly. For example, using the total number of agents in the simulation created, in figure 3.3 it takes 59 ticks for the message to spread around the network.

The next two snapshots show how a change in variable affects the time it takes for each message generated to spread in the network. At tick 79, in Figure 3.4, the message has spread the most it can as a result of agent acceptance or rejection, but is varied by having 880 nodes, 3 generators, 36 nodes that are likely to agree and 58 nodes that are likely to share.

When altering these variables in figure 3.4, the message spread ends at tick 178, signifying the end of this simulation run. In figure 3.5, the variables used are 500 nodes, 4 generators, 36 nodes that are likely to agree and 18 nodes that are likely to share.

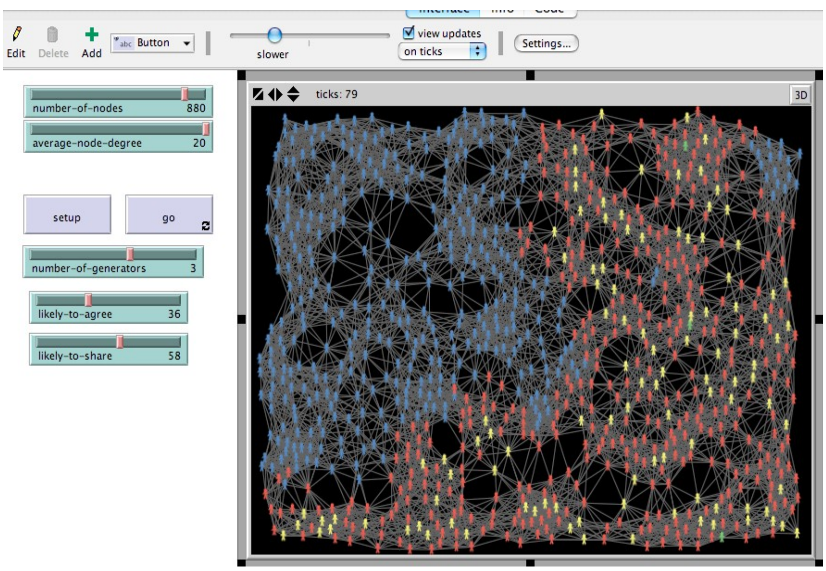

Figure 3.4: Simulation screen Shot D - virtual network of 880 nodes, showing the number of ticks at 79, with 36 nodes likely to agree and 58 nodes likely-to-share.

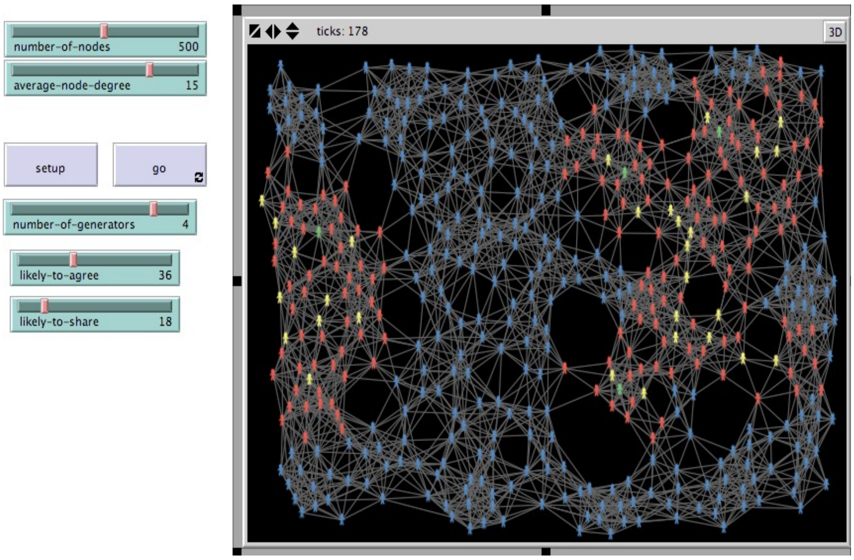

Figure 3.5: Simulation screen Shot E - virtual network of 500 nodes, showing the number of ticks at 178, with 36 nodes likely to agree and 18 nodes likely to share.

The two examples used in figures 3.4 and 3.5 show the time scale of how a message can spread in a network, altering agent behaviour that is mainly affected by the change in variables. In relation to how credible or malicious information is distributed across a network, it was anticipated that at data collection stages these variables would assist in the analysis process. This data was analysed and used to determine how a message is trusted with agents accepting or sharing such messages.

3.4.2 Preliminary findings

Understanding the changes in the patterns in the figures outlined was the most vital part of this research. It was anticipated that a basic comparison of figure 3.1 to figure 3.5 would most certainly raise questions whilst moving across each one. However, this only became possible as a result of the characteristics that agents were given. In figure 3.2 the changes that occurred when the "go" button in the simulation was clicked can be seen at tick 8; the bottom left corner, bottom right and top right corners displayed a change in colour codes. This signifies the spread of information with agents assuming different colours as a result of the characteristics they each carry.

At tick 79 in Figure 3.4, there was a clear change in data distribution when compared to figure 3.3. At this point the messages spread across the simulation assuming all bottom corners spreading to the top of the network before distribution stopped due to non-acceptance of the messages.

In figure 3.5, the parameters were altered and the resulting change is evident. Fewer nodes in the network at tick 178 (end of the simulation run) received the network and made decisions regarding what to do. Here the total number of nodes in the network was reduced to 500 with an average-node-degree of 15.

In the next phase of the simulation the researcher began to introduce characteristics into the network and monitor the agent's behaviour towards each one. It was anticipated that rewiring would be introduced at some point in the simulation process to measure how this practice alters an agent's response to individual characteristics. During this process data was collated for the research analysis.

3.5 Summary of this chapter

This chapter introduced what the term methodology is and went on to discuss two methods commonly used in research: qualitative and quantitative research methods. The choice of using the qualitative method was then justified, since the focus is on understanding and interpreting agent behaviour in a network. Next, which design and approach to be used was discussed, as well as whether the use of a network simulator would produce a virtual network with imaginary agents mirroring a real world network. To achieve this, the agent-based modelling (ABM) approach would be used, as it is a simulation approach spanning over 15 years in the industry and applied to various other research areas. This approach was then evaluated extensively using the ODD protocol for ABM's, which set the precedence for the design concept. ODD protocol is a guide set to assist a researcher to understand and create a virtual network that will answer its intended questions. Using this guide, the resulting factor was the design and creation of a preliminary model, with room for further updates, if required, throughout the experimental stages.

Chapter 4

The Experiment

4.0 Introduction

This chapter will discuss the following;

- The various processes carried out in the simulation and how the researcher achieved a functional simulation designed.
- The approach used for data generation and collection from the simulation based on the credibility hierarchy proposed in chapter one.
- A detailed discussion of how the experiment is expected to run showing how data is collected, including a description of the simulation with each run.

4.1 Experiment process

The researcher made use of Netlogo, a software used for agent-based modelling as discussed in chapter three. The model creates a virtual online social network with a particular number of nodes also called agents. After the design in Chapter 3 and the evaluation of the preliminary results; some updates were added to the model to mirror a real world network. Individual networks now have set down characteristics signifying their thoughts, beliefs or areas of interests in this case their area of interest. In this network, agents have been assigned with varied sets of these characteristics. During the simulation run, as agents interact with other agents, interests are subject to change as new information is received. Agents can pick up new interests adding to the list of interests they already hold or may drop some and in some cases stick to the interest they have without changing anything. This phenomenon will be consistent with real world networks and their behaviour.

We have classified agents into two types within our network: a message agent and a user agent (randomly assigned at different times). A message agent generates the information shared with other agents while a user agent is one that receives such information and propagates it in the network. The position of a message agent and a user agent is randomly assigned with each tick (a tick is the time it takes for a message to move from agent 1 to agent 2 within the network). So, an agent A could be a user agent at one point but changes to a message agent at a later time as a result of the random classification.

Characteristics or interest of agents in our experiment include; football, travel, lifestyle, rugby, swimming, politics, fashion, entertainment, health, tennis, science, technology, music, world (knowledge of the world), golf, skiing and education. Many of the agents have more than one preference meaning they have more than one characteristic quality. In this simulation agents have 1, 2 or 3 of these characteristics labelled as a preference. Also as the simulation runs, agents can change preferences as neighbouring agents sometimes influence their preferences.

Agents	Pref. 1	Pref. 2	Pref. 3
Person 3	Travel, Lifestyle	Politics	Skiing
Person 4	Tennis	Entertainment, swimming	Golf, health
Person 5	Swimming, Rugby	Golf	Tennis
Person 6	Lifestyle	Travel	Tennis
Person 7	Football, Music	Tennis	Rugby
Person 8	World, Golf	Health	Music, World
Person 9	Education, Music	Politics	Rugby
Person 10	Travel, education, tennis	Travel	Science

Table 4.1: Individual agent Preferences at Tick 1 randomly assigned.

Agents	Pref. 1	Pref. 2	Pref. 3
Person 3	Travel, Rugby, Lifestyle	Politics	Skiing
Person 4	Lifestyle, Travel Football, Tennis	Entertainment, swimming	Golf, Tennis, health
Person 5	Swimming, Rugby	World, Golf	Tennis
Person 6	**World, Lifestyle**	**Music, Travel**	**Tennis**
Person 7	Football	Tennis	Rugby
Person 8	World	Health	Music, Golf,World
Person 9	Education, Technology,	Politics	Rugby

	Music		
Person 10	Football, tennis	Fashion,Travel	Science
Person 11	Education	Science	Music
Person 12	Fashion	Education	Golf, World

Table 4.2: Agent Preferences at Tick 4 with changes in the preferences of person 6 (node 6).

The changes in preferences are in bold shown in table 4.2; person 6 seems to have taken on new preferences;

Person 6	Lifestyle	Travel	Tennis

Tick 1

Person 6	World, Lifestyle	Music, Travel	Tennis

Tick 4

These changes occur over time as the simulation runs.

4.2 The model

The model interface consists of a number of agents that are adjustable using a slider and number of messages also adjustable another slider. We have created two networks for the purpose of sampling Default network 1 and Default network 2. There is a go once button and a go button so the simulation can just carry on with go button and stops when there are no more changes or we use the go once to visually see changes at each tick. There is also a graph (message status) on the side of the interface to show how agents connect with messages generated by other agents in the network.

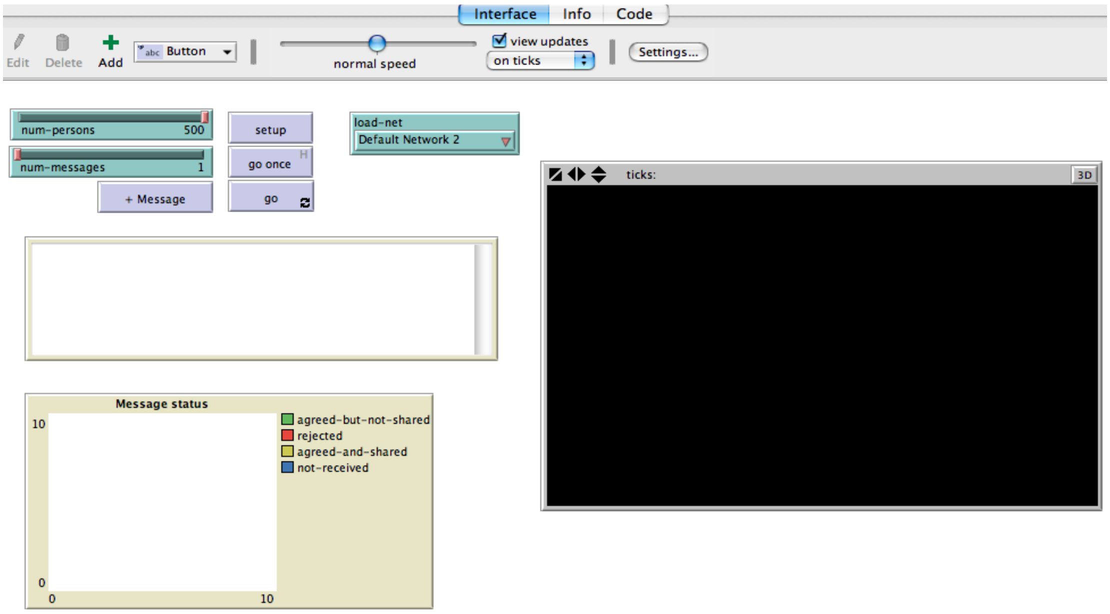

Figure. 4.1: Model showing the virtual network interface derived from the netlogo software

There are four categories of actions taken by agents;

- Agreed-but-not-shared (represented with the colour green); they are agents that have seen the message, trusted it but do not believe in it enough to share it.
- Rejected (represented with the colour Red); these agents have seen the message but have not trusted it therefore rejected it.
- Agreed-and-shared (represented with the colour yellow); these are agents that have seen the message, trusted it and have also shared it with their neighbours.
- Not-received (represented with the colour blue); these agents have not come across the message yet.

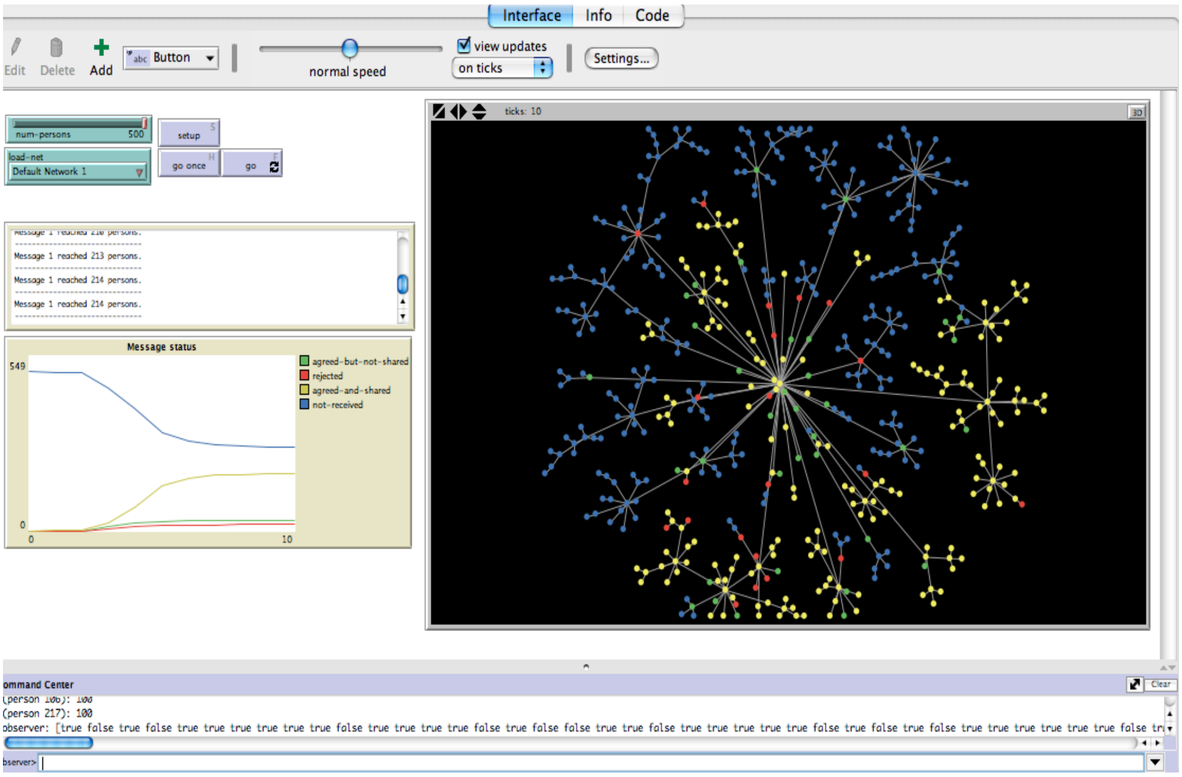

Figure. 4.2: Example of Model with activity showing nodes that have agreed-and-shared messages and some nodes that have not seen any message.

In figure 4.2, the plot illustrates the movements in the direction of acceptance of messages or rejected messages while some agents do not have a connection to the messages at this particular time. There is also a window showing what message has reached a number of agents so we have message 1 reaching 214 persons (agents) which means 214 agents came across/received the message.

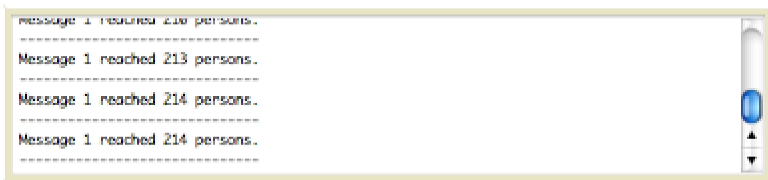

Figure 4.3 Screen shot of the output interface showing the number of nodes a message reached in this simulation run.

Once again we have adjusted the design of the model with the intent to mirror a real world network and record better data for our analysis. To begin with the initial status of all agents, at setup, all nodes that are not message generators remain blue. Agents that generate messages have been assigned colours based on validity of the message. In this network we have created 3 kinds of messages;

- Green (Message true)
- Purple (Message Suspicious)
- Pink (Message False)

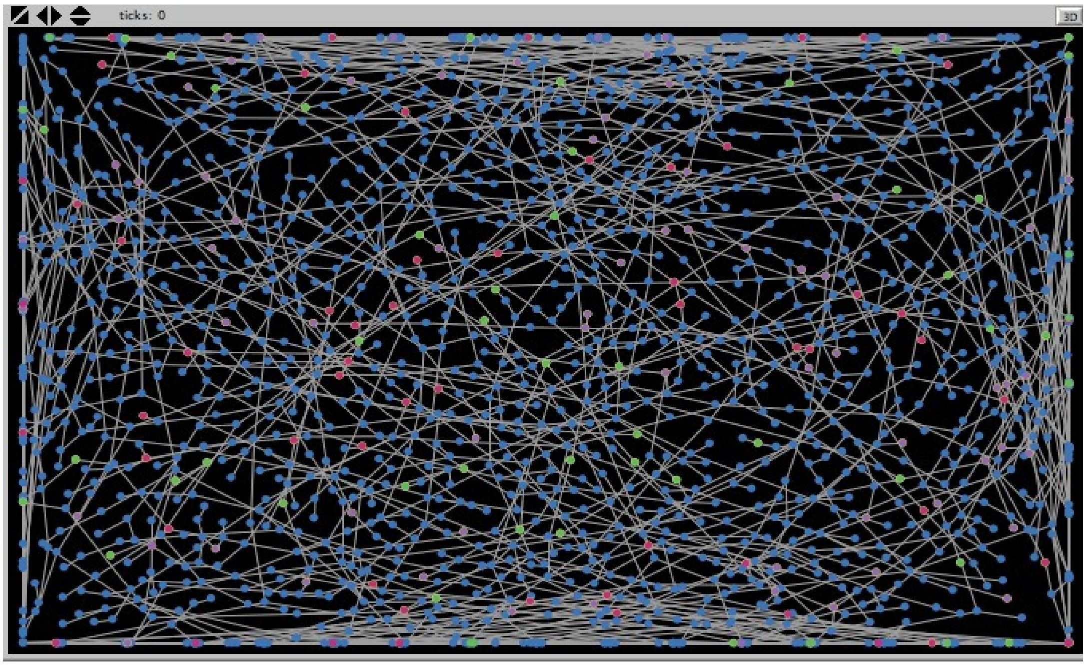

Figure. 4.4 Network at the setup stage showing a distribution of nodes with message generators.

The knowledge of whether the message is true, suspicious or false is not known to the agent. This is so that the agents can make their decisions independently based on their perception of a message and their neighbours deciding if to trust or not to trust. A probability calculation is added here in the code to produce a random decision for a given message. This calculation is derived using a modal system to calculate the number of neighbouring agents who have a preference relating to the message they have passed on thereafter make their decision.

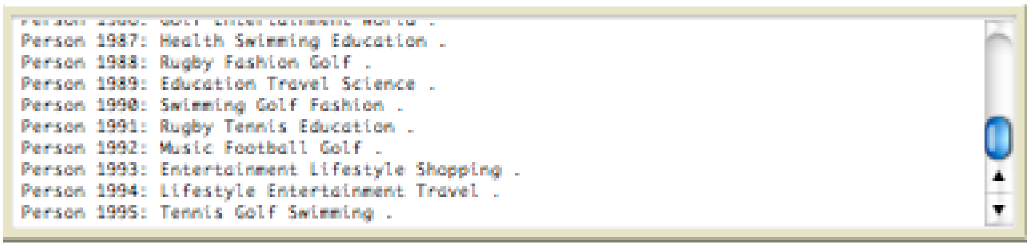

Fig. 4.4.1: Screen shot showing the preferences of agents according to each individual agent. Person 1989 has 3 preferences; education, travel and science.

Upon receiving a message, the following calculations will be used to record the reaction of nodes to messages.

If the person agent does not have the topic ID of the message that is received, it asks its neighbour's for opinion. Initially, the mean of scores was used to determine whether the person agent would share, accept or reject the message:

- Less than 70% is set to *reject*
- Greater than 70% but less than 100 is set to *accept*
- 100% is set to *share*

We put together possible patterns of message agent interaction with messages based on mean and mode calculations illustrated figure 4.4.2, 4.4.3 and 4.4.4

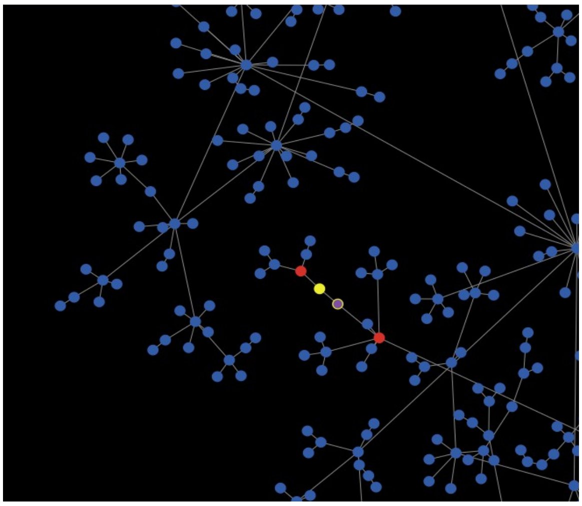

Figure 4.4.2: Screenshot with first mean calculation of agent reaction to messages.

Message 500 reached Person 317
Info on Person 317

- Preference = [12 14 3]
- Person 317 is directly connected to Person 371
- Person 317 is directly connected to person agents with ID number 393, 371 and 268.
- Among Person 317's direct links, Person 268 has the preference list [13 12 11]

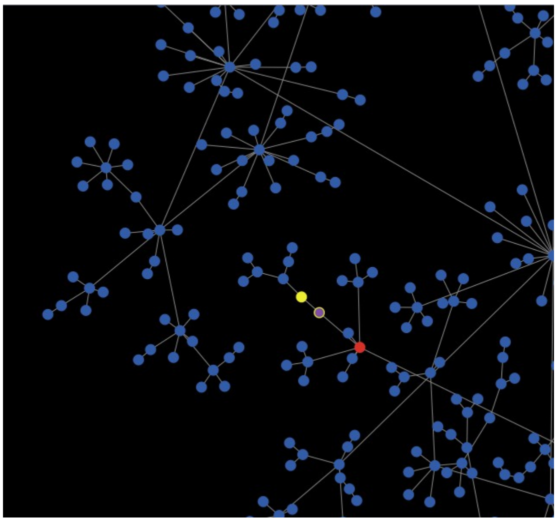

Figure 4.4.3: Screenshot with second mean calculation of agent reaction to messages.

Message 500 reached Person 371 and Person 402

Info on Person 371

- Preference = [4 3 17]
- Person 371 is directly connected to Person 172
- Person 371 trusted Message 500 because it's Person 172's 1st preference

Info on Person 402

- Preference info [12 16 2]
- Person 402 is directly connected to Person agents with ID number 172, 51, 302, 258, 177, 427 and 202.

An alternative is the use of **mode** in the command to retrieve the most common score or scores for a particular message, this is based on the frequency of a message. Figure 4.4.4 shows the sequence of changes based on the same setup used previously for the mean:

- Message 500 initiator is Person 172
- Message 500 topic is 11 (Rugby)

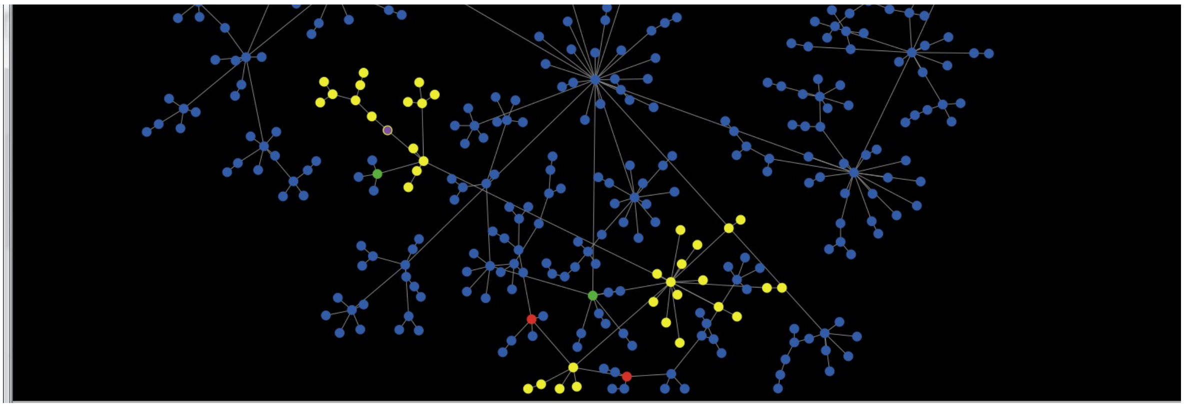

Figure 4.4.4: Screenshot with mode calculation of agent reaction to messages.

4.3 Simulation Process

When the setup button is clicked agents have to take one 1 of the 4 categories available to them as we see in the colour changes in the network.

It follows these steps;

1. The initiator sends a message (i.e. message 1 with topic technology) to its neighbours.
2. If the neighbour's preference list contains the topic id for technology, it automatically accepts it. At the same time may decide to share the message.
3. If the neighbour's preference list does not have the topic id for technology, it will randomly decide if it's going to accept it or not based on the acceptance-chance parameter, which is the mode. The number of persons around the agent that have accepted the message.
4. The neighbours' acts as initiator when sharing the message, thereby following the laid down guidelines while propagating the message.

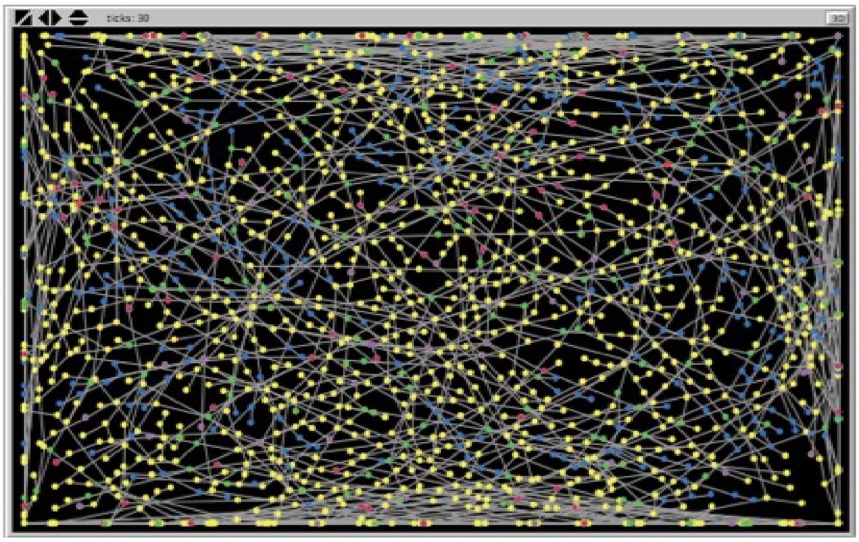

Figure. 4.5: Network interface showing activity after 1 simulation run

4.3.1. Results explained

In this network we have 1000 agents in constant communication with one another. When a message breaks out in the network through a single agent other agents begin to distribute the message passing it on to other agents in the network. In the simulation we have tasked 20 of these nodes with the initial outbreak of our message and they are to distribute this message to the other agents in the network. We have altered some of the messages carried by these agents to contain false and suspicious messages, as would normally be the case in the real world when a rumour surfaces online. When information is uploaded in any online OSN most agents (people) tend to rewrite the message to appeal more to their targeted audience making reference to past activities relating to the particular message. So they alter the original message by adding some photographic images to it to get the attention of other agents.

At each tick we have recorded the outcome of the nodes that have come across the message. In our first simulation the result recorded the total tick time was recorded at 19 and is represented by the table 4.3. A tick is the time it takes for a message to reach an agent in the simulation run time. This means at every tick we expect that an agent would have received the message from another agent depending on their preferences and connection with their neighbours. However, because of the characteristics set aside for each node having agents make the decision of accepting information or sharing them based on their preferences and that of their neighbours it is likely that if the agents do not trust the message they will not share it with other agents. We have also proposed some guidelines to identify the agents within the four stages of our hierarchy in the results produced in the simulation.

The guidelines are:

1. If a message is shared by 3 or more agents and has not been rejected by any agent that message could be seen as Experienced Credibility.
2. If we have 2 agents that agrees and not shared the message with 1 that has shared it that message could be seen as Surface credibility.
3. With 1 agent rejecting the information and 1 agreed not shared then the message could be seen as Reputed Credibility.
4. With 2 or more agents rejecting the message irrespective of what other neighbouring agents think this message is presumed credible.

An example of the result from the simulation is illustrated in Table 4.3. Here the data generated shows the acceptability, rejection and shared rate of the message in the network based on the proposed guidelines.

Accepted	Shared	Rejected	Hierarchy
3	9	1	Rep.
0	3	1	Rep.
1	2	1	Rep.
4	15	2	Pre.
3	8	4	Pre.
2	6	0	Exp.
10	67	4	Pre.
0	2	1	Rep.
1	2	1	Rep.

Table 4.3: Simulation result with Credibility Status proposed based on the proposed guidelines.

4.4 Summary of this chapter

In this chapter we have discussed the process of the experiment and how the model was designed. Using the ODD protocol a preliminary design a preliminary was created to begin a simulation process. To initiate the process agents, are given the task of message creation or generation, which is done by randomly assigning agents with this process. In the next stage each agent was randomly assigned characteristics and the researcher observed agents showing autonomous behaviour by taking up each characteristic independently over time. This behaviour likens this virtual network to a real world online social network where agents act independently. A description of the network model and the attributes that make up the model, which included a graphical representation as the model began each simulation run was analysed also in this chapter. Finally the researcher illustrated how the message sharing in the network is achieved as the simulation executes with data generated from the model recorded and interpreted.

Chapter 5

Results

5.0 Introduction

In this chapter we continue recording and evaluating the data generated after each simulation run. This is based on the number of agents in the network, their characteristics and the number of messages in circulation during the run. We begin by running the simulation with an initial network made up of 500 agents, then during the simulation run we begin reducing the number of messages introduced into the network. We then look out for changes or patterns that may arise as a result of the reduction in messages. This data is then recorded accordingly and analysed.

5.1 Simulation run 1 results

The simulation for this run has a network of 500 nodes and 100 message agents. Figure 5.1 shows a screenshot of the interface of the network and we see the agents colour coded based on the decision they have taken individually according to their preferences. The number of ticks is the time it has taken for 100 messages to be distributed in the network for this particular message and for this simulation the tick time shown in the screenshot is 26. It is important to note that not all nodes would have received a message at the end of each simulation run.

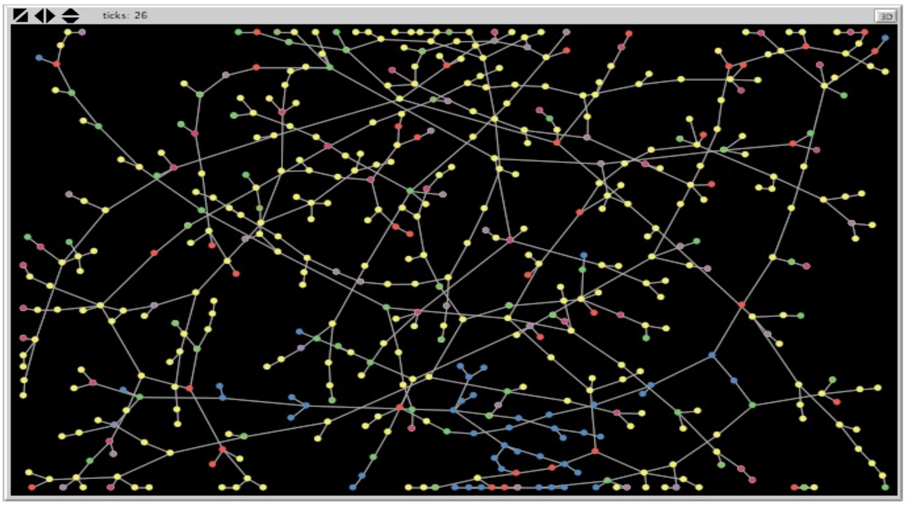

Figure. 5.1. Screenshot of network showing changes in a nodes decision as each one receives messages during a simulation run.

The result produced is given below with screenshot of a graph from Netlogo illustrating the movement of the messages in the network and how agents go about accepting or rejecting the message based on their preferences.

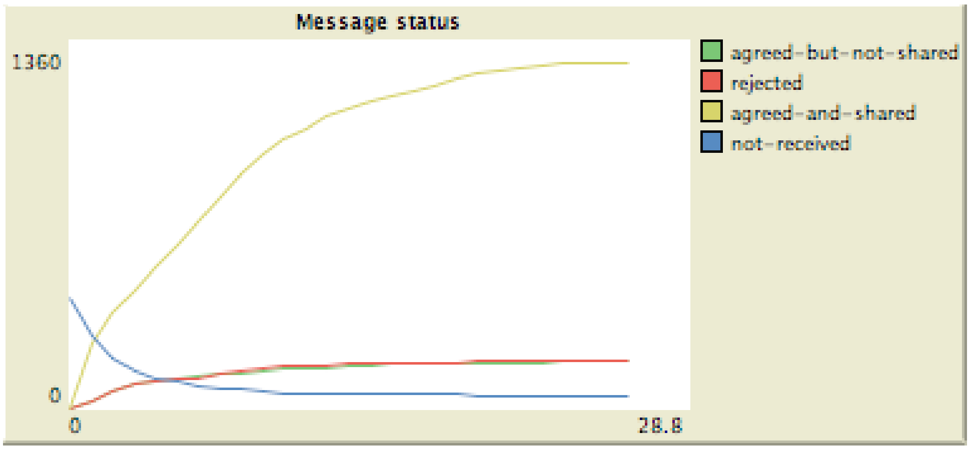

Figure. 5.1.1: Network graph of distribution of messages in the network

We illustrates figure 5.2 in table 5.1 to show how messages are distributed amongst nodes.

Time in ticks	Pn	Kn	Qn	Dn
0	0	0	0	408
1	25	27	27	272
2	68	62	62	187
3	95	95	95	143
4	101	104	104	115
5	111	112	112	100
6	118	115	115	88
7	126	128	128	79
8	133	143	143	71
9	138	150	150	64
10	146	156	156	60
11	149	159	159	58
12	152	161	161	58
13	159	165	165	55
14	160	168	168	55
15	164	168	168	55
16	165	171	171	55
17	166	171	171	55
18	166	171	171	52
19	169	176	176	50
20	170	178	178	48
21	170	178	178	47
22	170	178	178	46
23	173	179	179	45
24	173	179	179	45
25	173	179	179	45
26	173	179	179	45

Table 5.1: Data of simulation run 1 showing the number of nodes in each category.

Key

Pn = Agreed-not-shared

Kn = Rejected

Qn = Agreed and Shared

Dn = Not received

This table records the number of node changes over time (ticks). It shows the number of nodes that have received the messages based on four categories listed on the table; agreed not shared (PN), agreed and shared (QN), rejected (KN), not received (DN). With this data we have ascertained how many agents did not get the opportunity to make a decision to share or reject before the simulation run ended based on the sole fact that they did not receive the message.

The results from the numbers show a correlation between pair of columns such as;

1. Pn and Qn; as the messages are propagated in the network the number of nodes who agree but not shared (PN) and those who agree and share (QN) begin to increase between tick 0 and tick 26. Correlation between PN and QN = 0.997680221.

2. Pn and Dn; this correlation shows a reduction in the number of nodes that have not received a message (DN) as compared to those that have but not shared (PN) with other nodes. Correlation between Pn and Dn = -0.950959129.

3. Qn and DN; this correlation represents the number of nodes that agree and have shared (QN) compared to before the message was received (DN) between tick 0 and tick 26. Correlation between Qn and Dn = -0946215449.

An illustration is shown in the corresponding graphs for column Pn, Qn and Dn.

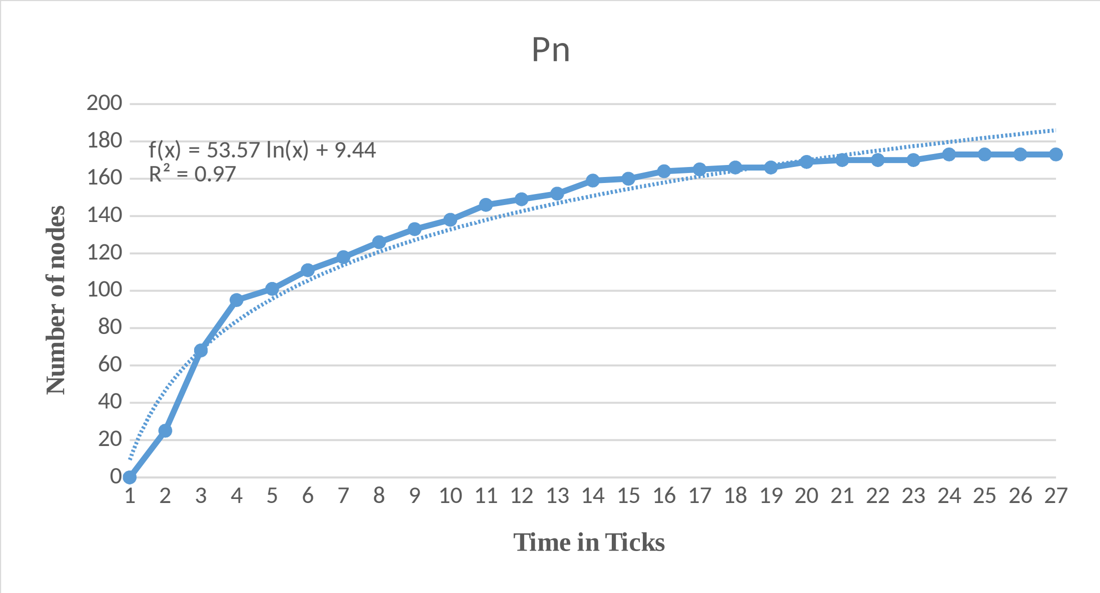

Figure 5.2: Graph of Pn (agreed not shared). The y-axis is the node changes while the x-axis is the number of ticks (the time it takes for a complete simulation run).

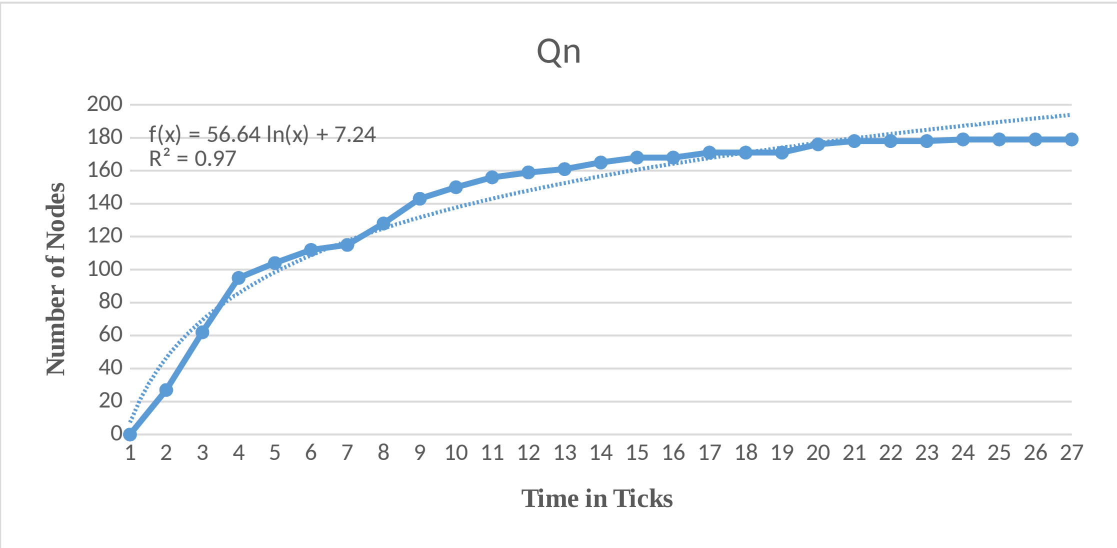

Figure 5.2.1: Graph of Qn (agreed and shared). The y-axis is the node changes while the x-axis is the number of ticks (the time it takes for a complete simulation run).

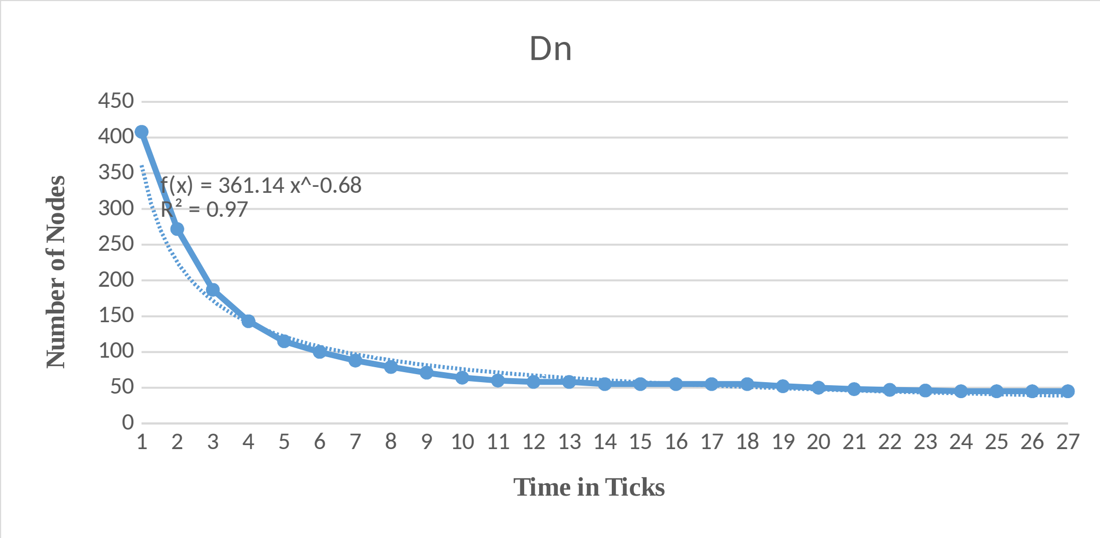

Figure 5.2.2: Graph of Dn (not received). The y-axis is the node changes while the x-axis is the number of ticks (the time it takes for a complete simulation run).

5.2 Simulation run 2 results

Number of messages in this run has been reduced to 95 but the total number of agents in the network was held at 500. This is to enable the researcher evaluate what changes will occur with a reduced number of messages.

Time-n/ticks	Pn	Kn	Qn	dn
0	0	0	0	415
1	19	30	228	286
2	65	64	349	198
3	88	83	447	153
4	95	92	533	132
5	99	105	630	109
6	104	113	725	91
7	111	126	800	76
8	118	129	851	73
9	124	130	915	70
10	134	135	971	67
11	139	138	1026	64
12	143	141	1078	61
13	150	144	1120	59
14	151	146	1150	58
15	154	152	1178	54
16	155	157	1200	49
17	157	160	1215	47
18	159	161	1225	44
19	161	163	1235	44
20	162	163	1242	44
21	163	163	1249	44
22	164	164	1254	44
23	164	164	1258	44
24	165	165	1263	44
25	165	165	1274	44
26	165	166	1282	44
27	165	166	1287	44
28	165	166	1291	44
29	165	166	1296	44
30	165	166	1299	44
31	165	166	1306	42
32	165	167	1313	42
33	165	167	1321	40
34	168	167	1326	39
35	168	167	1334	39
36	168	167	1339	39
37	168	168	1344	39
38	168	168	1353	39
39	168	168	1359	39
40	169	168	1363	39
41	169	168	1365	39
42	169	1 68	1366	39
43	169	168	1366	39

Table 5.2: Data of simulation run 2 showing tick times and number of nodes that have made a decision based on message received.

The time taken (number of ticks) here is 43 more than the previous simulation whose time was 26 ticks. This here shows that the more the messages in a network of 500 agents the less time it take for agents to come in contact with the message. This also means the decision to accept, reject or share is done faster and agents get acquainted with their neighbours giving room for a wider range of opinion formation based on the messages received within the network. However the patterns for accept, share or reject is not that different as in the first simulation run.

The results also show a correlation between the following pair of columns;

1. Pn and Qn at 43 ticks correlation = 0984691.
2. Pn and Dn at 43 ticks correlation = -094679.
3. Qn and Dn at 43 ticks correlation = -0.90034.

Graph's for each column is illustrated thus;

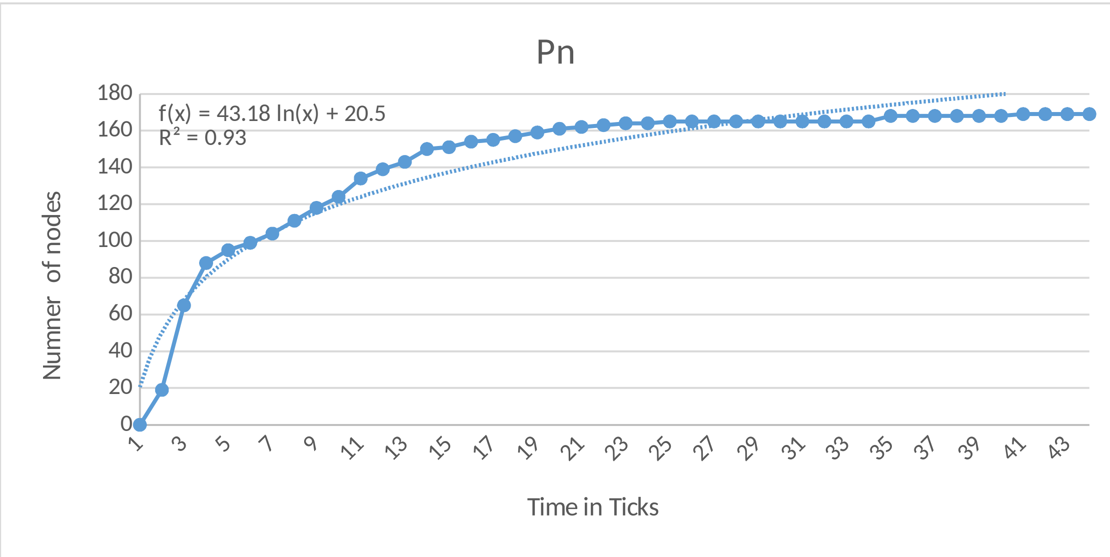

Fig. 5.3: Graph of Pn (agreed not shared). The y-axis is the node changes while the x-axis is the number of ticks (the time it takes for a complete simulation run).

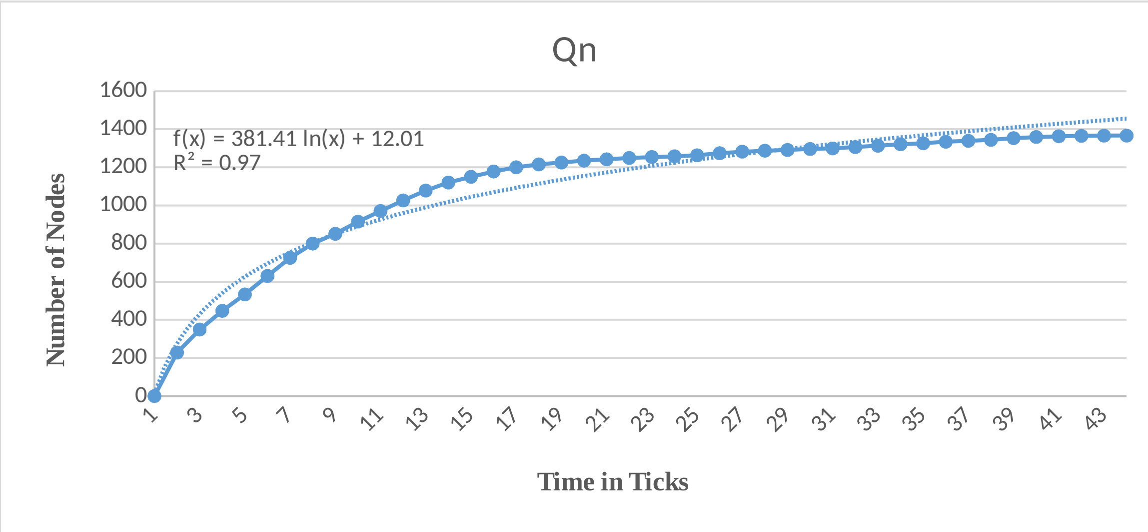

Fig 5.3.1: Graph of Qn (agreed and shared). The y-axis is the node changes while the x-axis is the number of ticks (the time it takes for a complete simulation run).

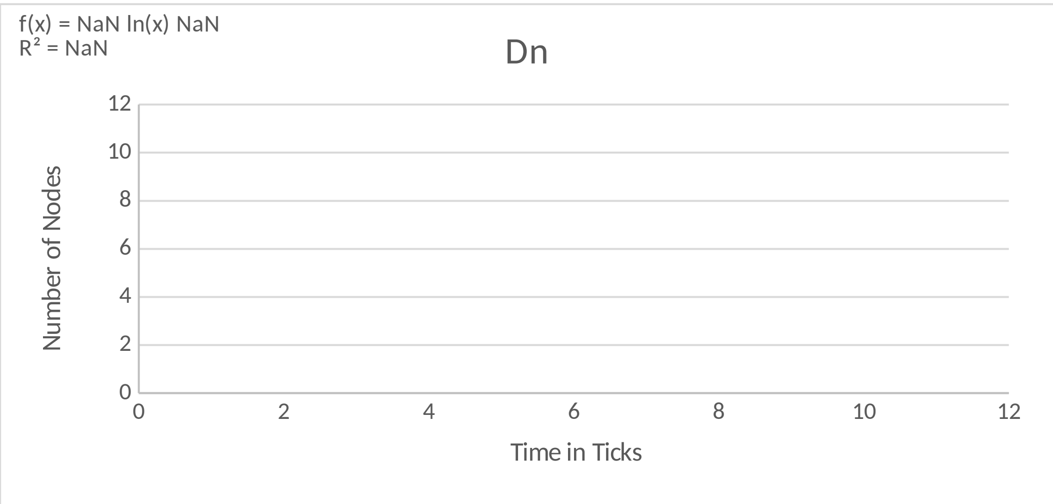

Fig 5.3.2: Graph of Dn (not recieved). The y-axis is the node changes while the x-axis is the number of ticks (the time it takes for a complete simulation run).

5.3 Simulation run 3 results.

The number of messages for the third simulation was reduced to 80 and the time it took for a complete run was 15 ticks.

Time= n/ticks	Pn	Kn	Qn	dn
0	0	0	0	427
1	25	22	201	312
2	52	47	291	237
3	75	79	354	183
4	80	94	407	160
5	84	97	451	140
6	93	101	486	129
7	96	104	509	122
8	99	106	525	115
9	101	111	538	109
10	102	113	549	107
11	102	113	555	105
12	106	113	561	103
13	107	114	564	102
14	108	114	564	102
15	108	114	564	102

Table 5.3 Data from simulation run 3

The data in table 5.3 did not follow the pattern of the preceding 2 simulation runs as was anticipated, rather the messages were propagated in less time however the correlations and graphs were not too different.

1. Pn and Qn at 15 ticks correlation = 0.991717.
2. Pn and Dn at 43 ticks correlation = -099172.
3. Qn and Dn at 43 ticks correlation = -0.98989.

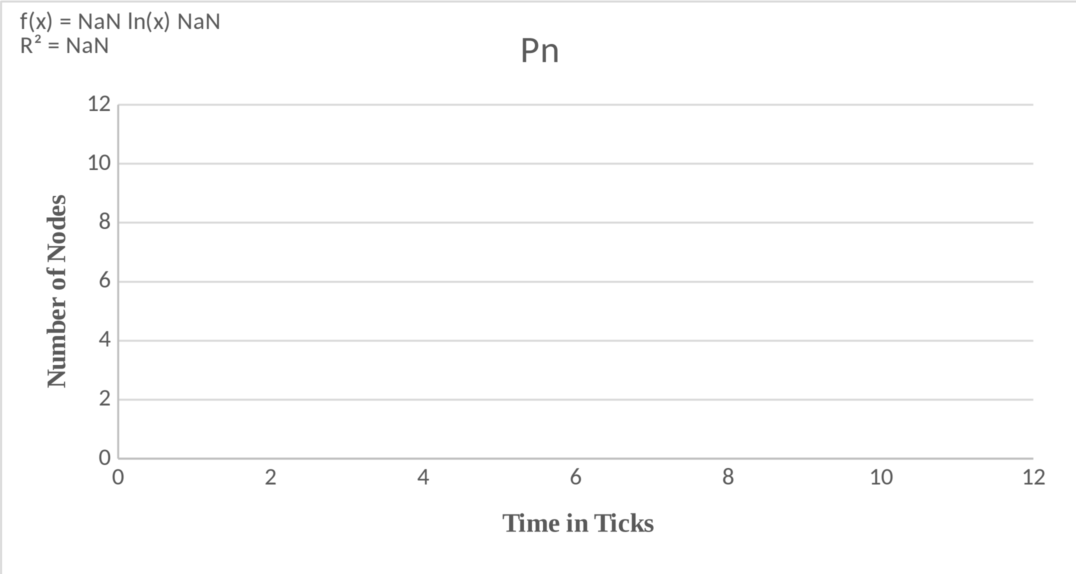

Figure5.4: Graph of Pn (agreed not shared). The y-axis is the node changes while the x-axis is the number of ticks (the time it takes for a complete simulation run).

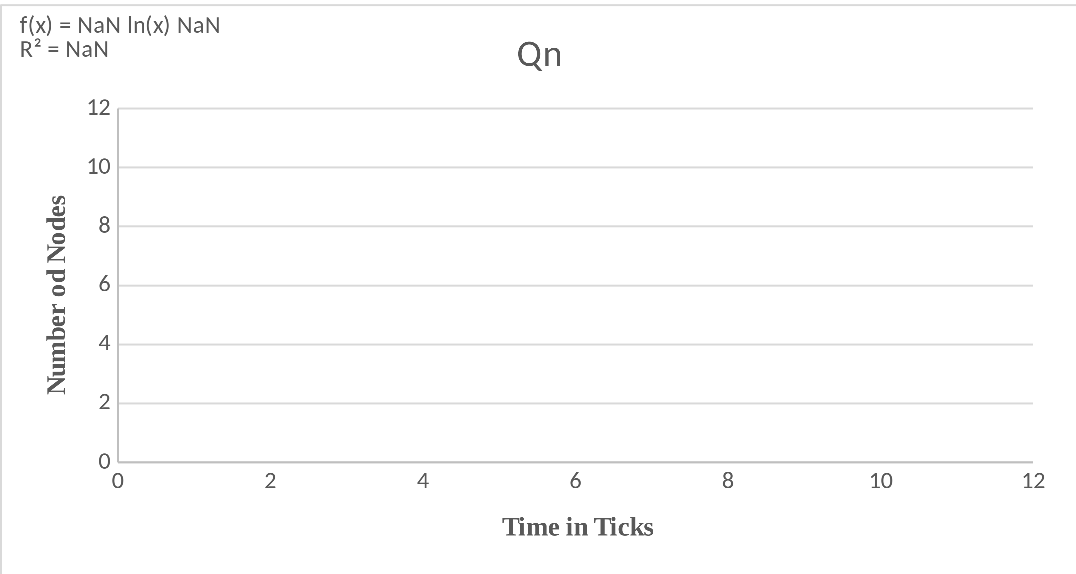

Figure 5.4.1: Graph of Qn (agreed and shared). The y-axis is the node changes while the x-axis is the number of ticks (the time it takes for a complete simulation run).

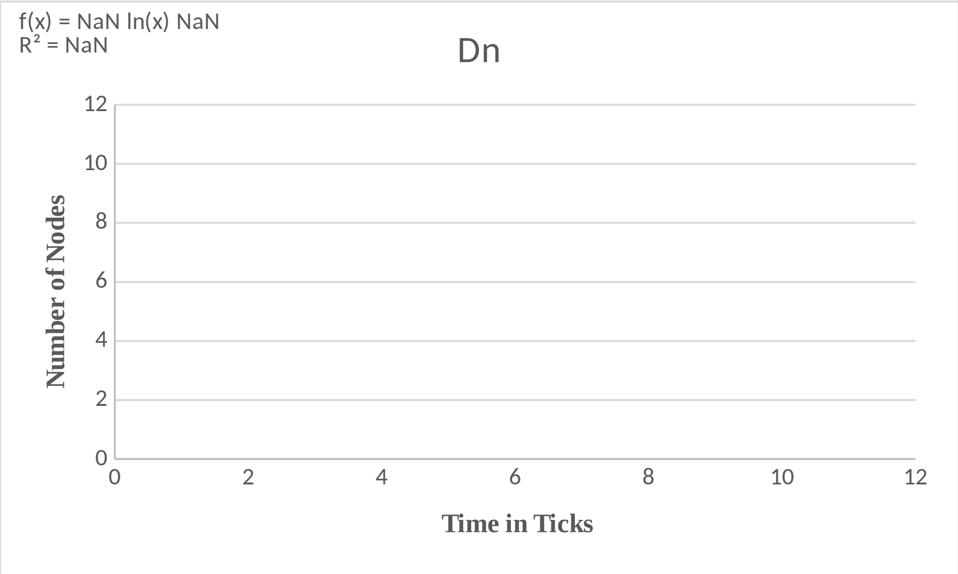

Figure 5.4.2: Graph of Dn (not recieved). The y-axis is the node changes while the x-axis is the number of ticks (the time it takes for a complete simulation run).

The results from simulation run 1 to 3 shows agent behaviour as we alter the number of messages in the network at any one time.

5.4 CSV file results

Let us now review the CSV file generated from the simulation showcasing what number of nodes have accepted, shared or rejected the message. The CSV file also shows the specific node interest.

Message II	Info	Topic	Accepted	Shared	Rejected	Person Age	Person Age	Person Age
1	TRUE	Golf	0	2	1	[]	[209 7]	[442]
2	FALSE	Technolog	1	3	2	[392]	[413 363 1	[292 85]
3	TRUE	Technolog	2	6	2	[88 152]	[32 327 11	[353 27]
4	TRUE	Travel	1	3	0	[381]	[348 499 3	[]
5	TRUE	Lifestyle	0	1	2	[]	[6]	[316 490]
6	Neither Tr	Education	2	26	5	[14 183]	[70 155 77	[118 9 489
7	FALSE	World	2	13	1	[292 268]	[323 89 39	[480]
8	Neither Tr	Technolog	2	3	0	[453 286]	[109 255 4	[]
9	FALSE	Lifestyle	0	10	2	[]	[247 380 4	[236 212]
10	TRUE	Rugby	1	14	3	[212]	[104 222 4	[246 204 2
11	Neither Tr	Education	1	21	1	[292]	[174 168 3	[467]
12	Neither Tr	Music	1	1	0	[148]	[83]	[]
13	TRUE	Fashion	1	1	2	[294]	[420]	[23 148]
14	Neither Tr	Swimming	3	5	0	[261 367 8	[232 15 44	[]
15	TRUE	Golf	3	13	0	[84 115 30	[274 280 2	[]
16	FALSE	World	0	1	1	[]	[425]	[148]
17	FALSE	Science	2	9	0	[93 124]	[31 6 352	[]
18	TRUE	Music	5	71	4	[169 321 8	[470 156 2	[408 240 1
19	TRUE	Golf	1	3	0	[386]	[110 128 7	[]
20	Neither Tr	World	1	1	1	[139]	[246]	[267]
21	TRUE	Politics	1	2	0	[283]	[369 229]	[]
22	TRUE	Health	0	2	3	[]	[410 315]	[289 281 2
23	TRUE	Fashion	2	2	0	[361 360]	[367 269]	[]
24	TRUE	Music	0	10	2	[]	[9 216 256	[159 432]
25	Neither Tr	Skiing	4	38	5	[109 448 4	[108 275 2	[16 97 225

Table 5.4: CSV result generated showing the message status (true or false), the topic of the messages propagated and the number of accepted shared or rejected messages.

In the first column we have the messages listed 1 to 500 but only 25 is shown in our screenshot. The next column is the information tab that lets us know if the message is true, false or neither true nor false so we can say it's the status of the message. For example, from our screenshot message 6 is neither true nor false. The message status is not revealed to the other agents this is so that the agents can decide independently what credible and what is not which is the experience with real world social networks.

In the next column we have the topic of interest and in this instance message 1 topic of interest is golf. This status in this column is randomly assigned each time the simulation runs. In the next three columns we have the result of agent behaviour based on the message received (Accepted, Shared and Rejected). For message 6 in our screenshot, the message was accepted by 2 agents and from the 2 agents it was distributed to and by other agents in the network 26 times but 5 agents rejected the message. The remaining three columns list the specific agent number that accepted, shared or rejected the message, it is a long list of agent but is not required for the understanding of what the table represents.

5.4.1 Demonstrating the result; credibility hierarchy

The CSV file from simulation run 3 is used to establish the credibility hierarchy of individual nodes. We have indicated in the table below Experienced, Surface, Reputed and Presumed Credibility accordingly. Agents are positioned in the hierarchy according to the credibility framework proposed to see how many agents will be placed under each category. A reminder of the framework;

1. If a message is shared by 3 or more agents and has not been rejected by any agent that message could be seen as Experienced Credibility.
2. If we have 2 agents that agree and have not shared the message with 1 that has shared it, that message could be seen as Surface credibility.
3. With 1 agent rejecting the information and 1 agreed not shared then the message could be seen as Reputed Credibility.

4. If 2 or more agents reject the message irrespective of what other neighbouring agents think this message is presumed credible.

Accepted	Shared	Rejected	Hierarchy
3	9	1	Rep.
0	3	1	Rep.
1	2	1	Rep.
4	15	2	Pre.
3	8	4	Pre.
2	6	0	Exp.
10	67	4	Pre.
0	2	1	Rep.
1	2	1	Rep.
2	7	2	Pre.
1	5	1	Rep.
1	5	3	Pre.
0	1	1	Rep.
2	2	1	Rep.
1	10	0	Exp.
4	10	2	Pre.
5	22	3	Pre.
3	17	3	Pre.
0	1	1	Rep.
2	3	0	exp.
1	9	2	Pre.
0	5	1	Rep.
1	2	0	Sur.
1	2	0	Sur.

Table 5.5: Credibility status of nodes in CSV result

At the end of various runs and recording the csv result files, the tentative proposal for ranking messages according to the number of accepted, shared and rejected messages has proven tentatively effective with the anticipated result.

5.5 Summary of this chapter

This chapter sees series of result processes taken to achieve our final anticipated result. The generation of the CSV file shows the activities of the nodes, the frequency of interaction and how nodes react to messages. Messages can be accepted, shared or rejected as the CSV result shows. The resulting data from simulation run 1 were analysed with a graph thereafter, a calculation of the connections (correlations) that a column would have with another column; correlation between Pn (agreed-not-shared) and Qn (agreed-and-shared). Simulation run 2 was also analysed using this steps and so was simulation run 3. They each showed a consistency in the graph pattern and the connection of each column. In this instance it was a positive correlation result. The proposed framework is then introduced to attempt the classification of messages based on the behaviour of the nodes. As a result in table 5.5 we observed a number of messages categorised into the 4 types of credibility in the framework proposed.

CHAPTER 6

Validation

6.0 Introduction

This chapter aims to verify the results obtained in chapter 5 by using a real world Online Social Network data from twitter. We would discuss the following;

1. Twitter data analysis
2. The tool used for Visual analysis

3. Compare the twitter data with results of Netlogo simulation Data.

4. Evaluate the result of the analysis

5. How does the result compare to our Hierarchy theory.

6.1. Twitter data

Twitter is an online social networking site that allows users to communicate with other users. Nodes send out messages called Tweets, these tweets are read by other nodes and can be distributed in the network by sharing. The process of sharing messages is called re-tweeting. A message can be re-tweeted as many times as possible. The record for most re-tweeted message ever on twitter is held by Ellen DeGeneres; a Selfie (picture) taken at the Oscars in 2014 (twitter.com, 2016) was re-tweeted over 300 thousand times and counting.

Twitter has over 1 billion active users monthly with a record 500 million tweets sent out per day according to the last statistics (Twitter, 2016). In the research carried out by Kwak, *et al* (2010), it;

> *"classified the trending topics based on the active period and the tweets and it showed that the majority (over 85%) of topics are headline news or persistent news in nature"*.

Most news articles are tweeted a lot of the time before they even hit other media outlets; radio, TV and newspapers. The current trend shows celebrities, big corporations and even world leaders tweeting information directly to the rest of the world. They use this medium to share their thoughts on worldwide issues and make their own announcements when necessary. The traditional media seem to have lost the monopoly it once enjoyed as many world leaders and governing bodies are now using micro-blogging service like twitter (Solis, 2011). They tweet information and questions to their citizens and the rest of the world facilitating the distribution of information around the world in seconds (Solis, 2011). United states ex-president: President Obama was first to use Twitter during his campaign of 2008 @BarackObama. He is currently the most followed on twitter of all world leaders with over 85million followers (twitter.com@BarackObama, 2017). After the Tsunami of 2011 the Japanese government turned to twitter to provide updates as to the situation of the Tohuku-Pacific Ocean Earthquake; putting to rest false rumours (Solis, 2011). Tweets and re-tweets are seen and shared with millions of people daily which means the possibility of inaccurate information been re-tweeted is one that is inevitable. When nodes receive tweets the way they react to the message can shape their decision making on other tweets that the same node may send out in the future. Future tweets can be re-tweeted, challenged or ignored.

This real world phenomenon makes a strong argument for validating our virtual research data (Netlogo Data) with data generated from twitter. These networks have a large number of edges between nodes randomly assigned which would require tools to understand their complex and unpredictable nature. Such tools will aid in the extraction of data from the network, the visualisation and reporting of the network findings. There is a list of available software for network visualisation, some examples include but are not limited to; Inflow, Pajek, Cytoscope and Gephi. Many of the tools available are open source and can run on Windows, Linux and Mac OS X. In our research we have used Gephi to analyse our network data. Gephi was chosen because it is not only open source or requires little or no programing skills but also supports various types of networks (directed, undirected, mixed graphs and many more). It runs on multiple platforms, supports many network layouts, has the ability to visualise network in real time, exports reports to multiple formats (e.g., PDF, PNG, SVG, etc.). It does have its limitations; it is limited to visualising a network of only 50 thousand nodes and 1 million edges. This limitation does not affect this research, as we have not used a large data set of up to 50 thousand nodes. A desired reason for using Gephi is that it does not require a lot of expertise to understand it.

6.2 Acquisition of Twitter data using Twitter API and JSON Tool

6.2.1 Twitter API and JSON

Credibility has been researched in the past using Twitter as a case study. The research was focused on automatic methods of assessing the credibility of sets of tweets using analysis of microblog postings of trending topics to determine the credibility of tweets (Castillo *et al*, 2011). To acquire data from Twitter one would use the tool provided by Twitter and public data on this social network can be accessed using its Application Programing Interface (API). "An API is defined as a way for a program to accomplish a task, usually by retrieving or modifying data" (Twitter Inc. 2016). The API method allows programmers to make applications, widgets and projects that interact with twitter. Twitter offers two key elements for acquisition of its data: the Search API and the Streaming API (Highfield *et al*, 2013).

1. The search API is used to retrieve past tweets based on certain criteria's (for example; user, keyword, location, etc.) carried out within set limits. The search API returns limited number of tweets (messages) and therefore cannot be used to retrieve a complete archive of past tweets (messages) containing for example specific keywords. This is as a result of its in-built limits, that limits how many keywords or users that can be queried at any given time or within timeframes.
2. The streaming API on the other hand can be used to subscribe to a current and on-going stream of new tweets (messages) that contain specific keywords or originating from specific users or their location. As with the search API, there are significant

> limits on the number of users (nodes) or keywords that can be queried for data as well.
>
> (Highfield *et al*, 2013)

It is important to note here that some of the limitations of both elements provided by Twitter can be overcome through available third parties that resell twitter content online which are provided at a cost to the public (Highfield *et al*, 2013).

Given these limitation of the API, any research method which seeks to establish a realistic complete dataset of tweets (messages) related to specific keywords will need to begin tracking the keyword as soon as it first appears to avoid the risk of losing tweets (messages) as they will no longer be accessible using the search API. In addition, future messages must be recorded either by using the Streaming API to subscribe to an on-going update or by frequently retrieving the past tweets using the Search API (Highfield *et al*, 2013).

"*Even such retrieval methods cannot guarantee a comprehensive capture of Twitter data, however: outages on the side of server or client, or transmission problems between them, cannot be ruled out altogether, and may result in message loss*" (Highfield *et al*, 2013). The only means of actually crosschecking the dataset for its accuracy is via the API itself especially since the API constitutes the single window to access Twitter streams available to researchers. "*No dataset captured by using the Twitter API is guaranteed to be entirely comprehensive especially where research focuses on identifying broad patterns in Twitter activity from a large dataset. However, such research nonetheless remains valid and important*" (Highfield *et al*, 2013).

The data acquired from twitter in this research was done using the search API. The data pool was saved to JSON a software tool adopted to make the information readable.

JSON is a software that stores information in an organised and readable format. It is a lightweight data-interchange format and is easy for humans to read and write and also easy for machines to parse and generate. Based on a JavaScript Programming Language JSON is a text format that is completely language independent but uses conventions that are familiar to programmers of the C-family of languages. Ranging from Python, to Java, JavaScript, C, C++, C# and many others. Properties like this make JSON an ideal data-interchange language (JSON, 2016)

By using Jason the twitter data could now be understood so that further analysis could be carried out.

6.2.2 Extraction of Twitter data for this research

Extracting the path of any tweet is not a process provided by twitter API. The only possible option available is to scan a potentially huge number of users to isolate the path of some selected tweets. To extract the different paths would require computational power. This would mean that we need a much bigger pool of users to fully trace a re-tweet so that we can find the full path. It might be challenging, complicated and a time consuming process due to limitations of Twitter API. As a result the researcher decided to observe the single connections between users (based on re-tweets), although this is a simplified approach it should equally provide an insight into sharing (re-tweeting) behaviour. The lack of this feature meant we could only isolate a topic (example, music blogger) and for each node identified other nodes whose tweets were re-tweeted by the node.

We then sorted the data in a CSV file format that could then be analysed using Gephi. We sort to look at the number of times a message is re-tweeted by nodes, which translates to how frequently a message would have been shared. We anticipate that this would show the change in behaviour of other neighbouring nodes that could most likely affect the deciding factor for individual nodes when judging messages.

User ID/Tweet	**Retweet 1**	**Retweet 2**	**Retweet3**	**Retweet4**
TOIfanatics	fotografie	-	-	-
ADeerAHorse	vomitfaceband	NYDailyNews	getinthesea	nosuchpunk
maxfrost	GraceKuf	MoJamMondays	HeatherOnDrums	ScarletParke
ADMGPromotion3	-	-	-	-
Vivamagazines	NoelGallagher	RadioX	Marr_cissist	travellingsimon
SparkNShineRes	silvertorches	TheMaldivesBand	Do206	fshrmansvillage
fshrmansvillage	Tacoma416	StubbornSon	artisthome	dawniellerene
A_LoneArtOrig	-	-	-	-
ThatTacomaGirl	GrandCinema	-	-	-
HDrockelman	tadtheapp	intlsongcomp	davekusek	

TheAmbientLigt	dba256	TrueGritRecs	weareVIS	badformca

Table 6.1: Data recorded from Twitter extracted and displayed in a table

Table 6.1 shows an example of the raw data acquired from twitter. Data was extracted into excel tables for analysis. In this table, each row in the CSV file refers to a user A (node) while the related entries correspond to the user (other users) names (whose tweets were re-tweeted) at different times by user A.

For example; Timberfest have re-tweeted messages from @Tacoma416, @StubbornSon, @artisborne, and @dawnielerene. This means Timberfest has close connections with these users and would normally distribute the messages posted by these users. Resulting in some level of trust between Timberfest and the other users.

When we observe another node ThatTacomaGirl, there seem to be only 1 re-tweet of the message posted by 1 other node @GrandCinema. This would demonstrate that even though @ThatTacomaGirl is a part of the wider network it trust just 1 single node at this time to re-tweet its message on the network or has not seen the other messages or has decided not to re-tweet any nodes message at this time.

6.3 GEPHI

Gephi is an open source software available online for graph and network analysis. It makes use of a 3D render engine to present an enormous network in real time and to speed up the exploration (Bastian *et al*, 2009). Gephi is a flexible and a multi-tasking architecture that gives possible ways of working with complex sets of data producing valuable visual results. This piece of software not only shows a visualization of large graphs but one can import, filter, manipulate and export all types of networks (Bastian et al, 2009). The version used in our research is 0.9.1 (Gephi.org, 2016).

The graph section is seen in the middle, on the left hand side is a section for choosing the appearance of the graph (Attributes/Layouts). While on the right hand we have the context (counts of nodes and edges), various filters (Attributes, operator etc.) and statistics (Average node degree, graph density etc.).

6.3.1 Comparison: Simulation data and Twitter data using Gephi

Simulation data uploaded to Gephi.

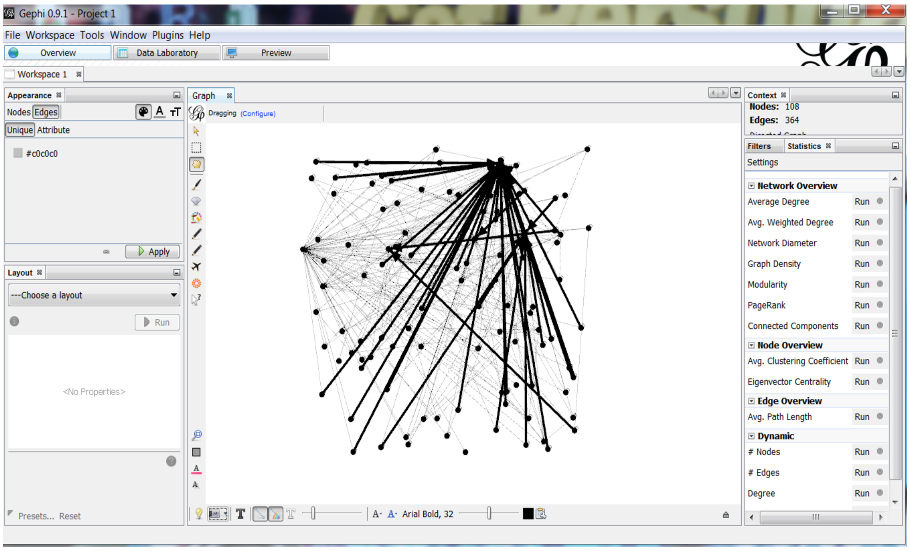

Figure 6.1: Gephi working interface with the research simulation data

The dataset used in this screenshot is that recorded from the research virtual Network, there are 108 nodes and 364 edges in this network. This is recorded in the context bar seen on the right hand side of the interface.

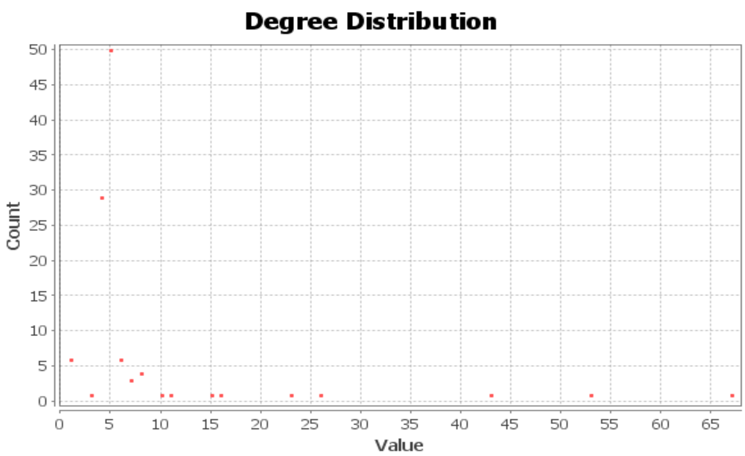

Figure 6.2: Average node degree simulated data

An average degree shows the number of edges linked to each node. This graph shows the nodes in a scale free state suggesting it randomly makes decision on messages based on the same network principles; opinion formation either as an individual or group of nodes to make a decision on a message.

Twitter Data illustrated using Gephi.

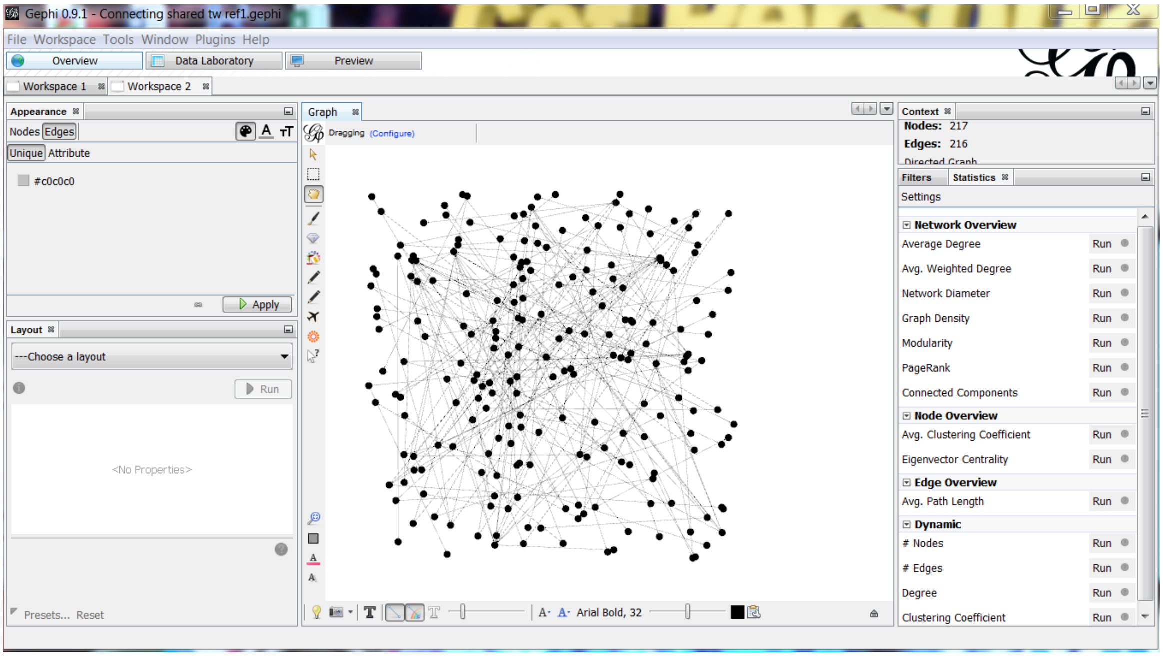

Fig. 6.3: Screenshot of Gephi interface with data from Twitter.

The dataset used in this screenshot is that recorded from Twitter, there are 217 nodes and 216 edges in this network. This is recorded in the context bar seen on the right hand side of the interface.

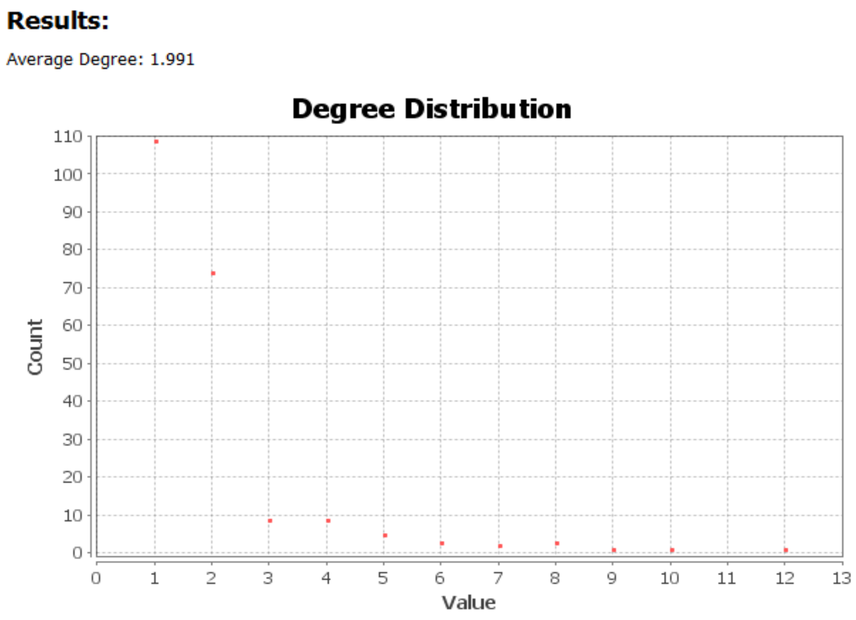

Figure 6.4: Average node degree twitter data

An average degree shows the number of edges linked to each node. In this data set the average number of edges individual nodes have is calculated at 1.991, illustrating that each node is connected to more than one node at a given time.

6.3.2 Illustrating credibility hierarchy with Netlogo data

In a Simulation run with 100 messages randomly distributed (only 15 messages represented in table 6.2; See Appendix Results Run 1 for full table) we categorise the number of accepted, shared and rejected according to the proposed rules for determining agent message credibility.

Message ID	Accepted	Shared	Rejected
1	0	2	1
2	1	3	0
3	3	44	2
4	2	1	0
5	1	9	2
6	2	12	4
7	0	1	1
8	2	4	1
9	0	5	2
10	6	116	12
11	1	28	6
12	5	50	4
13	1	6	2
14	0	1	1
15	2	18	2

Table 6.2: Simulation Results for run 1.

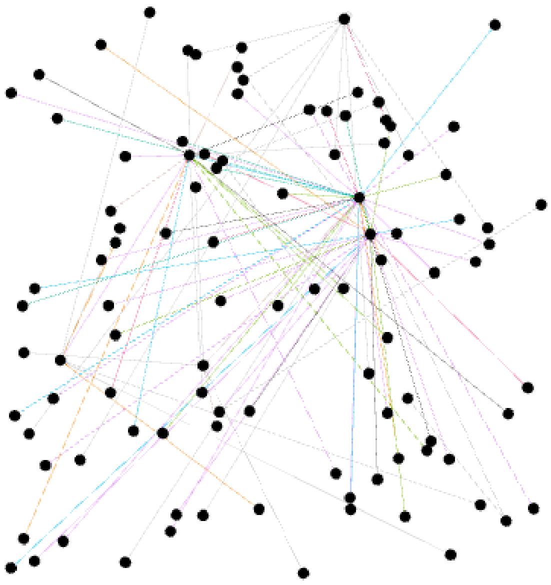

Figure 6.5: Simulation run 1 extract; 101 nodes and 100 edges

This figure shows a direct graph that has 101 nodes and 100 edges connecting each node with another node. What this graph means is that every node in the network is connected to at least one other node linking the node to others that are within its network. Some nodes may have multiple connections, which increases their chances of receiving new information in the network faster than nodes that are only connected to one node. So it is a case of the more connections you have the more likely you are to come across new or trending information.

6.3.3 Discussion: Data from virtual network and Twitter.

6.3.3.1 Netlogo data from simulation run 1

In Message ID 2, we have 1 accepted, 3 shared and 0 rejected. In this case, the message with ID number 2 has been shared 3 times with other nodes. This shared message would then either be shared or rejected based on how the next node perceives the message. Therefore, at this point the message was accepted by one node that then shared the message with 3 other nodes but no node rejected this message at this time. Going by the proposed framework this would classify this message on the experienced credibility section of our proposed pyramid.

6.3.3.2 Twitter data

In validating the twitter data with the data gathered from our Netlogo network, we try to match the data columns to that of data from Netlogo.

Represented here;

Message ID	**Source**	**Target**	**Shared (re-tweets)**
1	1	StuartMaconie	17
2	1	oliviaquillio	18
3	1	RidOfMeBlog	12
4	1	TheMusicGraph	13
5	1	TOIfanatics	1
6	1	ADeerAHorse	17
7	**1**	**maxfrost**	**5**
8	1	Vivamagazines	20
9	1	SparkNShineRecs	16
10	1	Timberfest	5

Table 6.3: Twitter data matched to resulting Netlogo table.

To compare this data let us place them side-by-side;

Message II	Accepted	Shared	Rejected
1	0	1	1
2	15	106	8
3	1	4	1
4	3	10	1
5	1	6	0
6	7	27	3
7	3	29	6
8	2	3	0
9	0	2	3
10	1	5	1
11	1	1	0
12	0	1	1
13	0	1	1
14	5	47	3
15	0	5	4

Table 6.4: Netlogo Data

ID	Source	Target	Shared
1	1	StuartMaconie	17
2	1	oliviaquillio	18
3	1	RidOfMeBlog	12
4	1	TheMusicGraph	13
5	1	TOIfanatics	1
6	1	ADeerAHorse	17
7	1	maxfrost	5
8	1	Vivamagazines	20
9	1	SparkNShineRecs	16
10	1	Timberfest	5
11	1	ThatTacomaGirl	1
12	1	HDrockelman	3
13	1	TheAmbientLight	11
14	1	seasyndrome	26
15	1	thefameriot	21

Table 6.5: Twitter Data

Firstly it is important to note that table 6.5 does not show any rejected data but by our assumption it is highly likely that some messages will be rejected even though that data is not represented in this table. In both tables we have an individual column showing shared message frequency. Let us now assume that the Source column in table 6.5 is the accepted column in table 6.4 and the shared column is same in both tables respectively. The ID column in 6.5 is the message topic at each time (in this case music bloggers) represented in 6.4 as the message ID (i.e the message topic earlier mentioned: music, fashion, sports etc).

Table 6.5 shows a single node re-tweeting messages from other nodes for a particular message this means the messages from these nodes are been shared.

We continue with our assumptions that for example; node “Maxfrost” shares messages of this topic Id 7, 5 times from 5 other nodes in the network showing his interest in the topic and the trust it has in these other nodes which makes it trust the messages they distribute.

These messages are most likely not going to be shared (re-tweeted) if Maxfrost has no prior relationship or trust for each of the originating nodes; this would therefore mean that over time Maxfrost could continue sharing messages from these nodes. This behaviour is seen in the over 100 nodes connections in the twitter data (see Appendix).

It has not been possible to classify twitter data with the credibility type hierarchy theory because of inadequate data availability from twitter during the acquisition process using twitter's API. The absence of the messages rejected by nodes or the pattern of reoccurring communication amongst nodes limits the accomplishment of this task at this time.

6.4 Summary of this chapter

The validation process was carried out using Twitter data. The similarities and frequency of the message distributions were observed. The researcher acquired data from Twitter using the API tool available via Twitter and it was tested against the design framework proposed. The validation process generated a set of behaviours that was different to the simulated data but there were similarities in the communication patterns of nodes in both networks. Messages where generated and distributed in the network by individual nodes and the frequency of message re-tweets showed the number of shares made for a particular message. It was not possible to highlight nodes that rejected messages due to the limitations in what data Twitter makes readily available to researchers. It is anticipated that were a full network data having these characteristic is fully assessable an analysis of both data could have yielded some of the desired outcome of tagging messages with individual credibility status.

It follows these simple steps;

1. The initiator sends a message (i.e. message 1 with topic technology) to its neighbours.
2. If the neighbour's preference list contains the topic id for technology, it automatically accepts it. At the same time may decide to share the message.
3. If the neighbour's preference list does not have the topic id for technology, it will randomly decide if it's going to accept it or not based on the acceptance-chance parameter, which is the mode. The number of persons around the agent that have accepted the message.
4. The neighbours' acts as initiator when sharing the message, thereby following the laid down guidelines while propagating the message.

Chapter 7

Evaluation, Limitations, Recommendation and Future Work.

7.0 Introduction

Growth and development of modern society hinges partly on information distribution. The Internet has created a social communication platform for millions of people around the world. Users of Online Social Networks take advantage of this service to communicate with friends and family whilst also getting updates on political and social news. A lot of data acquired from Online Social Networks are often times used and shared with other networks without any form of authentication process followed.

Most users tend to trust the information shared by a friend not questioning the credibility of the information, compared to one generated by an unknown user. A research into credibility was carried out on a larger scale, suggesting if trust could be placed in message distributed or in social network platform. This research outlined trustworthiness and expertise as factors of credibility that should be analysed. Twitter was also used in past research works as a case study into the analysis of credibility focusing on automatic methods of assessing the credibility of sets of tweets using analysis of microblog postings related to trending topics to determine the credibility of tweets.

However, this research was aimed at developing a framework that could assist in the assessment of the credibility of messages in Online Social Networks. We studied four types of credibility; experienced, surface, reputed and presumed credibility producing a credibility hierarchy. A virtual network was created using the agent-based modelling methodology to assess the credibility of messages distributed in the network to mirror a real world online social network. Nodes would generate messages randomly, passing the messages to other nodes and the behaviour of these other nodes towards the messages will then be recorded. The analysis of this data is carried out to determine the credibility of the messages. Some of the factors that were taken into consideration for the experiment were; Opinion formation, peer-to-peer networking, network rewiring and collaboration.

The analysis focussed on the behaviour of agents, the frequency of how nodes shared the messages that were distributed in the network. At the analysis stage, a framework was designed and the results from the analysis were tested using this framework.

7.1 Research Findings

The growing community of users in Online Social networks has in recent times made OSN's a new media outlet for breaking news around the world and there is a wide network of users within the wider network. These users (nodes) are interested in a range of topics or are professionals in the field; education, politics, entertainment, fashion or have no interest at all.

Nodes with no interest or no opinion rely on other nodes within the network to generate information and steer their belief in a message based on trust. This behaviour is typical in many online networks, as we have seen in our simulated network as nodes take on the characteristics of other nodes. The messages however are not evaluated or put through any kind of test that justifies any form of credibility. This formed the basis of this research work. Using Agent-based modelling to analyse the behaviour of nodes with various characteristics in a network.

This summary chapter will be discussed using the following headers;

1. The research approach carried out
2. The purpose for research approach
3. The contributions of the research approach to the research work
4. How each research objective has been addressed and the contribution of each one to the research work.

7.2 The research approach

This research work is centred on the how the behaviour of agents could or can be used in the assessment of credible messages propagated on the Internet through online social networks. An objective was created to quantify or qualify the desired goal of the research work, which the research would ultimately produce at the end of the research process. This goal would try to answer the research question proposed; can Agent-based modelling be a useful indication to assessing the credibility of messages in Online Social Networks.

7.2.1 Evaluation of this approach

In the literature review section we reviewed various topics and areas of interest relating to the research topic. These are possible attributes or behaviours individual nodes display in the network they are in at a given time. Some of the areas identified include; Peer-to-Peer networks, trust/credibility types, collaborative network and opinion formation. These are behaviours agent's display in many online networks and could assist in analysis of networks based on the desired result. This research has outlined the four types of credibility in a hierarchy with the intention to rank messages according to the individual characteristics of the agent responsible for their distribution within the network. Rewiring as an area of interest was also examined but more research will be carried on it in our future studies There has not been any reason for this rather the intention of the researcher was to analyse the first set of results from the network, thereafter adjust the network for rewiring, analyse its data then make a comparison of both data sets.

7.2.2 Methodology and rationale for choosing this method

The rationale behind choosing agent-based modelling was to record the behaviour of agents that would mirror that of those in a real world network. Using Netlogo for this service was to display the behaviour of nodes in a user interface that would be interpreted flexibly. The availability of sliders to adjust node numbers and/or messages for the observation of nodes behaviour at different simulation times was a great tool to have. Also it allowed data to be recorded with the writing of a code that automatically saved data during and after each run. This made it possible to achieve the aim of characterizing agents individually with random interest.

The results produced using this approach generated both qualitative and quantitative data and they each providing a useful insight into agent behaviour and the distribution of messages in Online Social networks.

7.2.3 Results of the mixed method

The results recorded in this research work shows both qualitative and quantitative data were generated. This mixed method has allowed the researcher to understand the behaviours of agents while at the same time having statistical figures of quantifying how many of these behaviours can affect the changes that occur in an Online Social network.

In chapter 5, table 5.4 shows a result of a simulated network having 500 nodes with 80 messages distributed in a simulation run. In this table record is taken of the following;

1. How many agents have accepted, shared or rejected the messages. This gives a statistics of the agent's decision based on the messages they receive each given time.
2. The table also shows the particular topic area individual messages carry, which would be a determining factor to the response of agents depending on their own areas of interest.
3. In this table we can also see the nature of a message; True, false and neither true nor false. This column shows the researcher the true nature of a message hence the resulting response of agent's decision to either accept, share or reject is understandable.

The columns with "Info and Topic" in Table 5.4 is essential in understanding this result because a vital part of the research focuses on agent behaviour towards the messages propagated in its network. So if we know that a message is true for example, but it is rejected by many nodes it could mean either that the message is been shared by an agent that has in the past shared a not so true message hence there is the suspicion that the current message could also not be true. Another meaning could be that it is relying on the opinion of neighbouring nodes towards the message before making its own decision. This assertion relates agent's behaviour in the simulation result to that of a real world network.

The mixed methodology used in this research work has been a single case study research. This is an approach used when analysing agent behaviour therefore it fulfils the research need we have carried out.

7.3 The purpose of this research approach

The mixed method approach was used or followed through as a result of the data that was generated both by the virtual network and the Twitter data collected. Qualitative methodology enabled the researcher to generate randomly assigned data based on agent behaviour, to explain and understand the characteristics that individual agents portray and the reaction of other agents to single agent behaviour. Quantitative allowed the research to record the figures of agents in the network and how their numbers would affect the behaviour of agents also. During the literature review we discussed the possible attributes the experiment design will follow and this yielded a mixed method of approach. We have highlighted some reasons why the method was used below;

1. The results from the simulation produced both qualitative and quantitative data; this further gives an understanding of agent behaviour and the characteristics that could be attributed to a message to show what credibility status would apply to it.
2. The procedure focussed on what method would be best in determining the end goal of the project. The use of agent-based modelling approach was to understand agent behaviour and how messages are distributed and perceived in a real world network. Netlogo was used to simulate a real world scenario of a network agent and in this case an Online Social network. The ODD protocol assisted in the experiment design taking into account the objectives the network would require in achieving the anticipated data that would be generated from the simulation. In understanding Netlogo it was

discovered that agents are identified using numbers, which means a recording of the quantity of agents, will be needed for during the analysis of the results.

3. After data collected from Netlogo was sorted into a CSV file. An observation of the graph produced by Netlogo during the simulation process shows the movements of messages and change in agent status. This shows the visual process of the network assisting in a better understanding of the network. This process outlines agent behaviour that shows the qualitative characteristics of the model also required for analysing the results.
4. To validate the Netlogo result we considered data generated by twitter. We have used Twitter data because it reflects the characteristics described in the data generated from our virtual network. Twitter as a network has Nodes (users), nodes generate messages and share them in the network while other nodes either re-tweet (share) the messages or simply do not based on who sent the tweet, re-tweeted, how they perceive the information, the opinion of their neighbours (in this case; friends). Twitter gives a platform for live updates as events unfold and is used worldwide. An API is also available through Twitter for collection of data by the public with some restrictions added. These attributes meet the criteria for using Twitter as a real world data for validating our virtual network. Twitter data showed the names of agents which we assumed our agent numbers while the re-tweet showed number of times other agents have shared the message. This also portrayed a mixed result of quantity and quality.
5. Analysis was carried out using both Netlogo and Gephi for simulated data and Gephi for Twitter data. A graphical representation of both data showed the interaction and communications carried out in the network. Gephi shows the average node degree in the network; the number of connections individual nodes has with other nodes in the

network. This indicates how the decision to accept, share or reject a message would have been measured.

All these procedures formed part of the objectives of the research (a key objective was to adopt an agent-based modelling method for understanding node behaviour) in an attempt to satisfy the research aim, a mixed method of approach was adopted.

7.4 Contribution of this approach to the research

The vast contribution of this approach can be seen in the analysis process. The use of agent-based modelling and its ODD protocol aided in formulating the concept for the simulation. This gave the initial framework of the experiment design and what the resulting outcome would achieve. Agent based modelling also allowed the researcher to produce in a network what seemed imaginary based on characteristics. The process of message generation and distributed in a network is a part of this creation process and is highlighted thus;

1. The initiator sends a message (i.e. message 1 with topic technology) to its neighbours.
2. If the neighbour's preference list contains the topic id for technology, it automatically accepts it. At the same time may decide to share the message.
3. If the neighbour's preference list does not have the topic id for technology, it will randomly decide if it's going to accept it or not based on the acceptance-chance parameter, which is the mode. The number of persons around the agent that have accepted the message.
4. The neighbours' acts as initiators when sharing the message, thereby following the rules and propagating the message.

These steps showed how the desired outcome of the simulation design was carried out. In chapter 5 we illustrated the status of messages according to characteristic of the agents carrying the message. This was ranked using a proposed test framework for determining if a message would be experienced, surface, reputed or presumed.

A major contribution of the approach used is in the discussion of agent connectivity in a network. The average number of connections between nodes in a network we assume shows that the decision making process for nodes could be determined by what neighbouring nodes are deciding or have decided. This is important because the message could be judged either based on a nodes perspective or on the opinions of other nodes.

7.5 Evaluation of each research objective

7.5.1 Critically review the function of online social networks, their agents (users) and how they operate

Reviewing online social networks has greatly impacted this research work because it has provided the knowledge needed to understand how these networks function. Firstly we focussed on what online social network provide as a service and this is the communication channel for friends and family but a commercial marketplace for businesses all over the world. News agencies have also taken advantage of the vast number of people that use this service to distribute breaking news stories to their audience in a quicker and efficient way. Individual users in online social networks are referred to as “agents”, they each carry out the basic function online social networks where supposedly invented for; creating and distributing information to other agents in their network. This could either be to family and friends or to acquaintances or strangers alike. Agents connect with other agents of similar or different opinions with regards to certain topics depending on goals. The combination of ideas could accomplish the goals they have set therefore agents sometimes engage in the act of rewiring. This is basically to bring in new agents into their network or disconnect with old or redundant agents. It is also the responsibility of all agents in a network to organize or reorganize the network except in a closed network system where only network administrators are authorized to do so.

This laid the foundation for the entire research; providing theoretical descriptions of activities carried out in online social network. Sufficient with this awareness, the researcher proceeded to the next objective.

7.5.2 Evaluating the Characteristics of agents and the credibility of the messages they generate

Agents are expected to demonstrate some degree of independence enabling them to fulfil their goals either independently or when working in a team. In the assessment stage of OSN's in the first objective we learned that agents have many characteristics. Also called software agents some of these characteristics include but are not limited to; autonomy, sociability, reliability, capacity to cooperate, ability to negotiate, sociability, flexibility, scalability and many more. These characteristics enhances their independent values and opportunity to carry out task without any form of interference however many agents even with their autonomic nature are sometimes easily influenced by other agents in their network. When this happens there maybe a change in their opinions towards a particular subject matter, which could in turn influence other agents in the network rightly or wrongly. Considering that many agents operate the majority of the time in teams, this could provide the platform for inaccurate message generation/distribution. To understand the credibility of messages four types of credibility were discussed; presumed, reputed, surface and experienced credibility. Based on the explanation of these four credibility types, the behaviour of agents and messages they share, there is a preliminary view of the type of credibility a message should be classed under.

This objective provides the means to understand agent characteristics and how they affect their behaviour in an online social network. By achieving this objective we were able to move on to the next objective; proposing a framework for the research was now a possibility.

7.5.3 Propose a draft framework to assess the behaviour of agents within an online network.

The behaviour of agents in online social network is connected with how agents perceive messages they come across that is; the messages that have been shared by other agents in their network. After evaluating trust and credibility and establishing four types of credibility the proposed draft framework places these credibility types in a pyramid (a hierarchy). So we begin with presumed credibility at the bottom, next at the top is reputed, then surface and experienced credibility placed at the very top of the pyramid. This signifying that messages propagated in a network that are presumed credible the researcher anticipates are less likely to be shared and could even be rejected by other agents also.

Here is the proposed framework;

1. If a message is shared by 3 or more agents and has not been rejected by any agent that message could be seen as having experienced credibility.
2. If we have 2 agents that agrees and not shared the message with 1 that has shared it that messaged could be taken to have surface credibility.
3. A message has reputed credibility when 1 agent rejects the message and 1 agreed not shared.

4. Lastly when 2 or more agents reject the message irrespective of the actions of other agents then that message is presumed credible.

The contribution of this objective to this research area is attributed to the detection of what messages fall under each of the four types of the credibility outlined. Also by understanding the behaviour of agents based on their characteristic it has become possible to predict what the resulting data from the model simulation would be. These characteristics would all be added to the model design aiming to answer the question of what makes a particular message credible.

7.5.4 Evaluate the credibility of messages passed between agents

To achieve this objective this research used a method known as the agent-based modelling method. Agent based modelling is used to design virtual networks for the purpose of generating data that aims to mirror real world networks. It is an established process (guidelines) used by many researchers to accomplish their research model designs. This guideline was used for the design of the preliminary model with Netlogo as the software of choice. Netlogo is software used for modelling abstract (virtual) networks; it is open source and cuts across operating systems. The network designed comprised of nodes and connecting edges with some agents tagged with the task of generating messages that will be shared and re-shared by other agents alike. In this model there were 1000 nodes and of this 1000 we had nodes that were message generators, nodes that have-not-seen the message, have-seen, have-seen-don’t-agree, have-seen-agree, have-seen-shared. We observed the model through a number of simulation runs to determine if it answered some questions geared towards achieving the aim of the research.

This objective enabled the detection of errors and possible areas to be corrected in the next phase of the model design. Re-running the simulation severally exposed many of the vulnerable and unproductive areas in the model and a plan to adjust these areas was taken on to the next phase.

7.5.5 Evaluate the draft framework and identify enhancement procedures of the model.

The simulation data recorded from the initial model design did not answer many the questions expected and from the records obtained after the initial model design some amendments were carried out in the model. Another model design was achieved after the enhancements; this model network had 500 nodes because when compared to the models that had 1000 or 5000 nodes, there were no significant changes in the results recorded. All nodes were task with the responsibility of generating new messages and sharing the messages to neighbouring nodes. Agents were also able to randomly take on new areas of interest and could be influenced by other nodes. All these made the agents behave like agents in a real world online social network would behave. The nodes had four categories; agreed-not-shared, rejected, agreed-and-shared and not-received. These categories of data were then used to determine which level in the pyramid messages belonged based on the framework proposed.

This objective answered many of the questions aimed at achieving the aim of this research. Its impact on the research is observed in the model's ability to generate data recorded from messages that have been shared, rejected or not seen. The data is analysed further using the framework proposed which establishes how agent credibility could be assessed.

7.6 Validating results with data acquired from Twitter; rationales for choosing twitter.

The decision to use twitter data in the validation of the research data was made because there were some network comparisons between both networks. Twitter is also scale free like the virtual network used for this research and has wide connections amongst nodes in the network, and only requires little or no knowledge of users (nodes) joining or leaving the network at random.

Data from Twitter was acquired using the 'twitter API', this data analysed in comparison with the data recorded from the virtual network result. There were some similarities and the behaviour most observed was the frequency of communication nodes have in the network. Nevertheless, available results were tested using the framework and success was recorded based on the concept of tagging the message status to the credibility hierarchy. Some recommendations have been made to this effect.

7.7 Limitations of the research

The limitations of the research were seen at the stage of validation, this is the stage when the researcher would verify the virtual network data with that obtained from Twitter. There were a couple of limitations and we have listed them thus;

1. Data for validation was not readily available because of the clauses Twitter holds. This limitation restricted some validation process for data at this time, which meant the researcher had to work with some theories. Number of shared messages where seen as messages tweeted and re-tweeted in twitter.
2. We observed during the validation process that the data recorded from Twitter did not correspondent completely with the simulated data from the research virtual network

because some of the required data from Twitter could not be acquired. Therefore the reflection of data using the framework designed could not be completely assessed however, some similarities of agents sharing information and re-tweets (in twitter) implied the data from the virtual network was similar to data obtained from twitter.

In the future we recommend where possible, a research on how more data from twitter that would include some rejected tweets (messages) can be assessed. This could be a tweet (message) that has been corrected by other agents because of its inaccuracy, which, could be highlighted as rejected.

7.8 Recommendations and future work

In The attempt to categorise nodes using the four types of credibility identified (experienced, surface, reputed and presumed credibility), this thesis has highlighted areas that will the basis for future research work.

These are some of the recommendations for future research work:

1. The model design in this research has not taken into account the presence of malicious agent that are specifically geared towards the disruption of the process of assessing the credibility of messages online so that is a recommendation that can be put into any future model. Recording this data will most likely alter the placement of messages in the hierarchy.
2. Evaluating rewiring procedure in detail, implementing the process in the model and observe the difference it would make in the network. One possible difference could be that nodes change their characteristics during every simulation run based on a general observation of all other nodes in the network. This may cause a change in decisions

made by other nodes also based on opinions from the general node population in the network. The other possible result could be that the characteristics of nodes change to resemble neighbouring nodes that could mean, all decisions are based on only neighbouring opinions. These decisions could be a deciding factor in how credible a message would be ranked in a network.

To reflect rewiring in our simulation the design will be altered to accommodate node connections and disconnections with other nodes randomly during a simulation run.

3. An enhancement to the proposed framework highlighting some specific message characteristics that would give messages a credibility status within the hierarchy. We would observe for certain characteristics that agents would possess to achieve an order of trustworthiness in the network. Messages generated or shared by these nodes will then attain a credibility status based on the decisions of nodes to trust either neighbouring nodes or the wider network of nodes.
4. The acquisition of data from sources like Facebook other than Twitter would be welcomed for a broader validation process and results findings.

Chapter 8

8.0 References

Akamine, S., Kawahara, D., Kato, Y., Nakagawa, T., Leon-Suematsu, Y.I., Kawada, T., Inui, K., Kurohashi, S. and Kidawara, Y., 2010, October. Organizing information on the web to support user judgments on information credibility. In *Universal Communication Symposium (IUCS), 2010 4th International* (pp. 123-130). IEEE.

Amoretti, M., Lafuente, A.L. and Sebastio, S., 2013, February. A cooperative approach for distributed task execution in autonomic clouds. In *2013 21st Euromicro International Conference on Parallel, Distributed, and Network-Based Processing* (pp. 274-281). IEEE.

Atzori, L., Iera, A., Morabito, G. and Nitti, M., 2012. The social internet of things (siot)–when social networks meet the internet of things: Concept, architecture and network characterization. *Computer Networks*, *56*(16), pp.3594-3608.

Barton, L. and Allan, V.H., 2007, November. Information Sharing in an Agent Organized Network. In *Proceedings of the 2007 IEEE/WIC/ACM International Conference on Intelligent Agent Technology* (pp. 89-92). IEEE Computer Society.

Barton, L. and Allan, V.H., 2008, December. Adapting to Changing Resource Requirements for Coalition Formation in Self-Organized Social Networks. In *Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT'08. IEEE/WIC/ACM International Conference on* (Vol. 2, pp. 282-285). IEEE.

Barabási, A.L. and Albert, R., 1999. Emergence of scaling in random networks. *Science*, *286*(5439), pp.509-512.

Bastian, M., Heymann, S. and Jacomy, M., 2009. Gephi: an open source software for exploring and manipulating networks. *International Conference on Web and Social Media (ICWSM)*, *8*, pp.361-362.

BBC (2011) *BBC 'must apologise' over Primark documentary*. Available at: http://www.bbc.co.uk/news/entertainment-arts-13794227 (Accessed: 7 November 2016).

Bentahar, J., Meyer, J.J.C. and Moulin, B., 2007, Securing agent-oriented systems: An argumentation and reputation-based approach. In *Information Technology, 2007. ITNG'07. Fourth International Conference on* (pp. 507-515).(2007) EEE.

Bian, B., 2012, May. Research of Factors on Impacting Internet Information Credibility Based on Electronic Commerce Users Demands. In *Communication Systems and Network Technologies (CSNT), 2012 International Conference on* (pp. 987-991). IEEE.

Bindra, G.S., Kandwal, K.K., Singh, P.K. and Khanna, S., 2012, July. Tracing Information Flow and Analyzing the Effects of Incomplete Data in Social Media. In *Computational Intelligence, Communication Systems and Networks (CICSyN), 2012 Fourth International Conference on* (pp. 235-240). IEEE.

Borgatti, S. (1998) *What is social network analysis?* Available at: http://www.analytictech.com/networks/whatis.htm (Accessed: 7 November 2016).

Bourdieu, P., 1990. *The logic of practice*. Stanford University Press.

Bradshaw, J.M., Carpenter, R., Cranfill, R., Jeffers, R., Poblete, L., Robinson, T., Sun, A., Gawdiak, Y., Bichindaritz, I. and Sullivan, K., 1997. Roles for agent technology in knowledge management: examples from applications in aerospace and medicine. In *Proceedings of the AAAI Spring Symposium, Artificial Intelligence in Knowledge Management*.

Brakmo, L.S. and Peterson, L.L., 1996. Experiences with network simulation. *ACM SIGMETRICS Performance Evaluation Review*, *24*(1), pp.80-90.

Brody, R., 2008, June. Credibility and information naiveté in socially networked information. In *2008 IEEE International Symposium on Technology and Society* (pp. 1-3). IEEE.

Canini, K.R., Suh, B. and Pirolli, P.L., 2011, October. Finding credible information sources in social networks based on content and social structure. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on* (pp. 1-8). IEEE.

Castillo, C., Mendoza, M. and Poblete, B., 2011, March. Information credibility on twitter. In *Proceedings of the 20th international conference on World Wide Web* (pp. 675-684). ACM.

Carrington, P.J., Scott, J. and Wasserman, S. eds., 2005. *Models and methods in social network analysis* (Vol. 28). Cambridge university press.

Chandrashekar, S. and Hockema, S., 2009, September. Online access, participation and information credibility assessment. In *Science and Technology for Humanity (TIC-STH), 2009 IEEE Toronto International Conference* (pp. 755-760). IEEE.

Cho, Y., Qu, G. and Wu, Y., 2012, May. Insider threats against trust mechanism with watchdog and defending approaches in wireless sensor networks. In *Security and privacy workshops (SPW), 2012 IEEE Symposium on* (pp. 134-141). IEEE.

Coen, M.H., 1991. Sodabot: A software agent construction system. *MIIT AI Laboratory, Cambridge, MA*.

Dickerson, M., 2014. Multi-agent simulation, netlogo, and the recruitment of computer science majors. *Journal of Computing Sciences in Colleges, 30*(1), pp.131-139.

Du, Y., Shi, Y. and Zhao, X., 2007, May. Using spam farm to boost PageRank. In *Proceedings of the 3rd international workshop on Adversarial information retrieval on the web* (pp. 29-36). ACM.

Edmonds, B. and Chattoe, E., 2005, July. When simple measures fail: Characterising social networks using simulation. In *Social Network Analysis: Advances and Empirical Applications Forum. Oxford, UK*.

Ellison, N.B., Steinfield, C. and Lampe, C., 2007. The benefits of Facebook “friends:” Social capital and college students’ use of online social network sites. *Journal of Computer-Mediated Communication*, *12*(4), pp.1143-1168.

Erd6s, P. and Rényi, A., 1960. On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci*, *5*, pp.17-61.

Fang, A., Wang, L., Zhao, J. and Wang, X., 2011, July. Social learning in multi-true-state networks. In *Control Conference (CCC), 2011 30th Chinese* (pp. 5784-5788). IEEE.

Fetterly, D., Manasse, M. and Najork, M., 2004, June. Spam, damn spam, and statistics: Using statistical analysis to locate spam web pages. In *Proceedings of the 7th International Workshop on the Web and Databases: colocated with ACM SIGMOD/PODS 2004* (pp. 1-6). ACM.

Fogg, B.J., Soohoo, C., Danielson, D.R., Marable, L., Stanford, J. and Tauber, E.R., 2003, June. How do users evaluate the credibility of Web sites?: a study with over 2,500 participants. In *Proceedings of the 2003 conference on Designing for user experiences* (pp. 1-15). ACM.

Fogg, B.J. and Tseng, H., 1999, May. The elements of computer credibility. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 80-87). ACM.

Furht, B. ed., 2010. *Handbook of social network technologies and applications*. Springer Science & Business Media.

Gephi.org. (2016) *Gephi - The Open Graph Viz Platform*. Available at: https://gephi.org [Accessed: 7 November. 2016).

Graham, K.C., 2010. Complexity science and social media: Network modeling in following "Tweets". In *2010 IEEE Systems and Information Engineering Design Symposium*.

Golbeck, J. and Parsia, B., 2006. Trust network-based filtering of aggregated claims. *International Journal of Metadata, Semantics and Ontologies*, *1*(1), pp.58-65.

Guimarães, S., Pereira, A., Silva, A. and Meira Jr, W., 2011, November. Credibility of web applications. In *Proceedings of the International Conference on Management of Emergent Digital EcoSystems* (pp. 25-32). ACM.

Guimaraes, S., Silva, A., Meira Jr, W. and Pereira, A., 2010, December. CredibilityRank: A Framework for the Design and Evaluation of Rank-based Credibility Models for Web Applications. In *Embedded and Ubiquitous Computing (EUC), 2010 IEEE/IFIP 8th International Conference on* (pp. 504-510). IEEE.

Gyongi, Z. and Garcia-Molina, H., 2005. Spam: It's not just for inboxes anymore. *Computer*, *38*(10), pp.28-34.

Gyöngyi, Z., Garcia-Molina, H. and Pedersen, J., 2004, August. Combating web spam with trustrank. In *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30* (pp. 576-587). VLDB Endowment.

Hashimoto, T., Kuboyama, T. and Shirota, Y., 2011, November. Rumor analysis framework in social media. In *TENCON 2011-2011 IEEE Region 10 Conference* (pp. 133-137). IEEE.

Hawe, P. and Ghali, L., 2008. Use of social network analysis to map the social relationships of staff and teachers at school. *Health Education Research*, *23*(1), pp.62-69.

Heidemann, J., Klier, M. and Probst, F., 2012. Online social networks: A survey of a global phenomenon. *Computer Networks*, *56*(18), pp.3866-3878.

Highfield, T., Harrington, S. and Bruns, A., 2013. Twitter as a technology for audiencing and fandom: The# Eurovision phenomenon. *Information, Communication & Society*, *16*(3), pp.315-339.

Hill, R., Antonopoulos, N. and Berry, S., 2012. Peer-to-peer networks and collective intelligence: the case for agency. *International Journal of Grid and Utility Computing*, *3*(4), pp.233-241.

Hjorth, A., Brady, C., Head, B., Wilensky, u., 2016. Turtles All the Way down: Presenting LevelSpace, a NetLogo Extension for Reasoning about Complex Connectedness. Construstionism, Bangkok, Thailand.

Hogg, T., 2009, November. Security challenges for reputation mechanisms using online social networks. In *Proceedings of the 2nd ACM workshop on Security and artificial intelligence* (pp. 31-34). ACM.

Investor.fb.com (2016) *Facebook - Facebook Reports Third Quarter 2016 Results*. Available at: https://investor.fb.com/investor-news/press-release-details/2016/Facebook-Reports-Third-Quarter-2016-Results/default.aspx (Accessed: 7 November 2016).

Ismail, S. and Ahmad, M.S., 2011, November. Emergence of social intelligence in social network: a quantitative analysis for agent-mediated PKM processes. In *Information Technology and Multimedia (ICIM), 2011 International Conference on* (pp. 1-7). IEEE.

Ismail, S. and Ahmad, M.S., 2011, February. Personal knowledge management among researchers: knowing the knowledge expert. In *Proceeding of the 10th International Research Conference on Quality, Innovation and Knowledge Management*.

Jennings, N.R., 2001. An agent-based approach for building complex software systems. *Communications of the ACM*, *44*(4), pp.35-41.

JSON (2016). *Introducing JSON*. Available at: http://www.json.org/ (Accessed: 7th Nov. 2016).

Kaiser, C., Krockel, J. and Bodendorf, F., 2010, August. Ant-based simulation of opinion spreading in online social networks. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on* (Vol. 1, pp. 537-540). IEEE.

Kaiser, C., Krockel, J. and Bodendorf, F., 2010, January. Swarm intelligence for analyzing opinions in online communities. In *System Sciences (HICSS), 2010 43rd Hawaii International Conference on* (pp. 1-9). IEEE.

Kaiser, C., Krockel, J. and Bodendorf, F., 2011, March. Analyzing Opinion Formation in Online Social Networks: Mining Services for Online Market Research. In *2011 Annual SRII Global Conference* (pp. 384-391). IEEE.

Kim, Y.A., Le, M.T., Lauw, H.W., Lim, E.P., Liu, H. and Srivastava, J., 2008, April. Building a web of trust without explicit trust ratings. In *Data Engineering Workshop, 2008. ICDEW 2008. IEEE 24th International Conference on* (pp. 531-536). IEEE.

Klein, D.B., 1997. *Reputation: Studies in the voluntary elicitation of good conduct*. University of Michigan Press.

Knocke, D. and Kuklinski, J.H., 1983. Network Analysis (Quantitative Applications in the Social Sciences).

Krebs, V. (2011) *Social network analysis: An introduction by Orgnet, LLC*. Available at: http://www.orgnet.com/sna.html (Accessed: 7 November 2016).

Ku, Y.C., Hsiao, H.W. and Wei, C.P., 2007, September. Analysis of web trust networks. In *Innovative Computing, Information and Control, 2007. ICICIC'07. Second International Conference on* (pp. 173-173). IEEE.

Kolbitsch, J. and Maurer, H.A., 2006. The Transformation of the Web: How Emerging Communities Shape the Information we consume. *J. UCS*, *12*(2), pp.187-213.

Kourtesis, D., Alvarez-Rodríguez, J.M. and Paraskakis, I., 2014. Semantic-based QoS management in cloud systems: current status and future challenges. *Future Generation Computer Systems*, *32*, pp.307-323.

Kwak, H., Lee, C., Park, H. and Moon, S., 2010, April. What is Twitter, a social network or a news media?. In *Proceedings of the 19th international conference on World Wide Web* (pp. 591-600). ACM.

Lee, J.D., 1991, September. The dynamics of trust in a supervisory control simulation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 35, No. 17, pp. 1228-1232). SAGE Publications.

Lengstorf, J. (2009) *JSON: What it is, how it works, & how to use it*. Available at: http://www.copterlabs.com/json-what-it-is-how-it-works-how-to-use-it/ (Accessed: 7th Nov. 2016).

Lewandowski, D., 2012. Credibility in Web search engines. *Online credibility and digital ethos: Evaluating computer-mediated communication*, pp.131-146.

Li, W., Wang, X. and Zhu, Q., 2013, August. Topology analysis and clustering for localized network in SINA Weibo. In *Smart and Sustainable City 2013 (ICSSC 2013), IET International Conference on* (pp. 321-324). IET.

LinkedIn (2016) *About LinkedIn*. Available at: https://press.linkedin.com/about-linkedin (Accessed: 7 November 2016).

Liu, B., Terlecky, P., Bar-Noy, A., Govindan, R., Neely, M.J. and Rawitz, D., 2012. Optimizing information credibility in social swarming applications. *IEEE Transactions on Parallel and Distributed Systems*, *23*(6), pp.1147-1158.

Liu, S., Ying, L. and Shakkottai, S., 2010, September. Influence maximization in social networks: An ising-model-based approach. In *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on* (pp. 570-576). IEEE.

Maes, P., 1994. Agents that reduce work and information overload. *Communications of the ACM*, *37*(7), pp.30-40.

Mansoury, M., Shahriari, H.R. and Shajari, M., 2012, May. Trust evaluation using items ratings in online communities. In *20th Iranian Conference on Electrical Engineering (ICEE2012)* (pp. 780-783). IEEE.

McNabb, D. E., 2010. *Research Methods for political Science: Quantitative and Qualitative Approaches*, M. E. Sharpe Inc.

Menges, F., Mishra, B. and Narzisi, G., 2008, December. Modeling and simulation of e-mail social networks: a new stochastic agent-based approach. In *Proceedings of the 40th Conference on Winter Simulation* (pp. 2792-2800). Winter Simulation Conference.

Mialhe, F., Becu, N. and Gunnell, Y., 2012. An agent-based model for analyzing land use dynamics in response to farmer behaviour and environmental change in the Pampanga delta (Philippines). *Agriculture, ecosystems & environment*, *161*, pp.55-69.

Miao, G., Tao, S., Cheng, W., Moulic, R., Moser, L.E., Lo, D. and Yan, X., 2012, April. Understanding task-driven information flow in collaborative networks. In *Proceedings of the 21st international conference on World Wide Web* (pp. 849-858). ACM.

Mitchell, B. (2016) *What's peer-to-peer software?* Available at: https://www.lifewire.com/p2p-networking-and-software-818019 (Accessed: 7 November 2016).

Muir, B.M., 1987. Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies*, *27*(5-6), pp.527-539.

Moro, G., Ouksel, A.M. and Sartori, C., 2002, July. Agents and peer-to-peer computing: a promising combination of paradigms. In *International Workshop on Agents and P2P Computing* (pp. 1-14). Springer Berlin Heidelberg.

Mu, B. and Yuan, S., 2010, June. A method for evaluating initial trust value of direct trust and recommender trust. In *Computer Design and Applications (ICCDA), 2010 International Conference on* (Vol. 2, pp. V2-185). IEEE.

Nogueira, D., Rocha, L., Santos, J., Araújo, P., Almeida, V. and Meira, W.J., 2002, November. A methodology for workload characterization of file-sharing peer-to-peer networks. In *Workload Characterization, 2002. WWC-5. 2002 IEEE International Workshop on* (pp. 118-126). IEEE.

Pastor-Satorras, R. and Vespignani, A., 2001. Epidemic dynamics and endemic states in complex networks. *Physical Review E*, *63*(6), p.066117.

Paprzycki, M. and Abraham, A., 2003. Agent systems today; methodological considerations. In *Proceedings of 2003 International Conference on Management of e-Commerce and e-Government* (pp. 416-421).

Pinheiro, C.A.R., 2011. *Social network analysis in telecommunications* (Vol. 37). John Wiley & Sons.

Pluchino, A., Garofalo, C., Inturri, G., Rapisarda, A. and Ignaccolo, M., 2014. Agent-based simulation of pedestrian behaviour in closed spaces: a museum case study. *Journal of Artificial Societies and Social Simulation*, *17*(1), p.16.

Porter, J. and Sen, S., 2007, November. Searching for Collaborators in Agent Networks. In *Proceedings of the 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology-Workshops* (pp. 508-511). IEEE Computer Society.

Probst, M.J., Park, J.C., Abraham, R. and Kasera, S.K., 2010, October. Socialswarm: Exploiting distance in social networks for collaborative flash file distribution. In *Network Protocols (ICNP), 2010 18th IEEE International Conference on* (pp. 263-274). IEEE.

Reaney, P. (2012) *Most of world interconnected through email and social media*. Available at: http://www.reuters.com/article/uk-socialmedia-online-poll-idUSLNE82Q02120120327 (Accessed: 7 November 2016).

Riegelsberger, J., Vasalou, A., Bonhard, P. and Adams, A., 2006, April. Reinventing trust, collaboration and compliance in social systems. In *CHI'06 Extended Abstracts on Human Factors in Computing Systems* (pp. 1687-1690). ACM.

Resnick, P., Kuwabara, K., Zeckhauser, R. and Friedman, E., 2000. Reputation systems. *Communications of the ACM*, *43*(12), pp.45-48.

Roussopoulus, M., Baker, M., Rosenthal, D., Guili, T., Maniatis, P. and Mogul, J., 2004, February. P2P or Not 2 P2P. In *The 3rd International Workshop on Peer-to-Peer Systems, IPTPS*.

Rowe, M., 2010, April. The credibility of digital identity information on the social web: a user study. In *Proceedings of the 4th workshop on Information credibility* (pp. 35-42). ACM.

Russell, S., Norvig, P. and Intelligence, A., 1995. A modern approach. *Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs*, *25*.

Ryan, P., 2011. *Digital and Information Literacy: Social Networking*. The Rosen Publishing Group, Inc.

Sewell Jr, W.H., 1992. A theory of structure: Duality, agency, and transformation. *American journal of sociology*, pp.1-29.

Scott, J., 1991. Social Network Analysis: A Handbook SAGE Publications Ltd. *London, UK*.

Sheridan, T.B., Vámos, T. and Aida, S., 1983. Adapting automation to man, culture and society. *Automatica*, *19*(6), pp.605-612.

Sherman, L. W. (2003) *SOCIOMETRY IN THE CLASSROOM*. Available at: http://www.users.miamioh.edu/shermalw/sociometryfiles/socio_are.htmlx (Accessed: 7 November 2016).

Sridhar, U. and Mandyam, S., 2011, July. Information sources driving social influences: a new model for belief learning in social networks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on* (pp. 321-327). IEEE.

Steinmetz, R. and Wehrle, K., 2004. Peer-to-peer-networking &-computing. *Informatik-Spektrum*, *27*(1), pp.51-54.

Suzuki, Y., 2010, November. A credibility assessment for message streams on microblogs. In *P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), 2010 International Conference on* (pp. 527-530). IEEE.

Talia, D., 2011, July. Cloud Computing and Software Agents: Towards Cloud Intelligent Services. In *WOA* (Vol. 11, pp. 2-6).

Tanaka, K., Ohshima, H., Jatowt, A., Nakamura, S., Yamamoto, Y., Sumiya, K., Lee, R., Kitayama, D., Yumoto, T., Kawai, Y. and Zhang, J., 2010, January. Evaluating credibility of web information. In *Proceedings of the 4th International Conference on Uniquitous Information Management and Communication* (p. 23). ACM.

Tran, H.T., Domerçant, J.C. and Mavris, D.N., 2016. A Network-based Cost Comparison of Resilient and Robust System-of-Systems. *Procedia Computer Science, 95*, pp.126-133.

Tseng, S. and Fogg, B.J., 1999. Credibility and computing technology. *Communications of the ACM*, *42*(5), pp.39-44.

Treepuech, W., 2011, December. The application of using social networking Sites with available online tools for teaching and learning management. In *IT in Medicine and Education (ITME), 2011 International Symposium on* (Vol. 1, pp. 326-330). IEEE.

Twitter (2016) *it's what's happening*. Available at: https://about.twitter.com/company (Accessed: 7 November 2016).

Twitter (2016) *The search API — Twitter developers*. Available at: https://dev.twitter.com/rest/public/search (Accessed: 7 November 2016).

Varien, M.D. and Potter, J.M. eds., 2008. *The social construction of communities: agency, structure, and identity in the Prehispanic Southwest*. Rowman Altamira.

Victor, P., Cornelis, C., De Cock, M. and Herrera-Viedma, E., 2011. Practical aggregation operators for gradual trust and distrust. *Fuzzy Sets and Systems*, *184*(1), pp.126-147.

Viejo, A. and Castellà-Roca, J., 2010. Using social networks to distort users' profiles generated by web search engines. *Computer Networks*, *54*(9), pp.1343-1357.

Villatoro, D., Sabater-Mir, J. and Sen, S., 2011, May. Social instruments for convention emergence. In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 3* (pp. 1161-1162). International Foundation for Autonomous Agents and Multiagent Systems.

Vorakulpipat, C., Marks, A., Rezgui, Y. and Siwamogsatham, S., 2011, July. Security and privacy issues in Social Networking sites from user's viewpoint. In *2011 Proceedings of PICMET'11: Technology Management in the Energy Smart World (PICMET)* (pp. 1-4). IEEE.

Wang, H., Wang, F. and Liu, J., 2012, March. Accelerating peer-to-peer file sharing with social relations: Potentials and challenges. In *INFOCOM, 2012 Proceedings IEEE* (pp. 2891-2895). IEEE.

Wang, J.C. and Chiu, C.C., 2008. Recommending trusted online auction sellers using social network analysis. *Expert Systems with Applications*, *34*(3), pp.1666-1679.

Wang, X.F. and Chen, G., 2003. Complex networks: small-world, scale-free and beyond. *IEEE circuits and systems magazine*, *3*(1), pp.6-20.

Wang, Y., Xing, L. and Wang, H., 2013, January. Reliability of scale-free complex networks. In *Reliability and Maintainability Symposium (RAMS), 2013 Proceedings-Annual*, pp.1-6. IEEE.

Watts, D.J. and Strogatz, S.H., 1998. Collective dynamics of ‘small-world’networks. *nature*, *393*(6684), pp.440-442.

Wasserman, S. and Faust, K., 1994. *Social network analysis: Methods and applications* (Vol. 8). Cambridge university press.

Weisberg, M., 2013. Modeling herding behavior and its risks. *Journal of Economic Methodology*, *20*(1), pp.6-18.

Weisberg, M., 2012. *Simulation and similarity: Using models to understand the world*. Oxford University Press.

Wilson, C., Boe, B., Sala, A., Puttaswamy, K.P. and Zhao, B.Y., 2009, April. User interactions in social networks and their implications. In *Proceedings of the 4th ACM European conference on Computer systems:* pp. 205-218. Acm.

Wilensky, U., 2014. Computational thinking through modeling and simulation. *Whitepaper presented at the summit on future directions in computer education. Orlando, FL. http://www. stanford. edu/* coopers/2013Summit/WilenskyUriNorthwestern REV. pdf*.

Wilensky, U. and Rand, W., 2015. *An introduction to agent-based modeling: modeling natural, social, and engineered complex systems with NetLogo*. MIT Press.

Wilensky, U., details and use, terms of (1999) *NetLogo home page*. Available at: https://ccl.northwestern.edu/netlogo (Accessed: 7 November 2016).

Witkowski, J., 2011, July. Trust mechanisms for online systems. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence* (Vol. 22, No. 3, p. 2866).

Wooldridge, M. and Jennings, N.R., 1995. Intelligent agents: Theory and practice. *The knowledge engineering review*, *10*(02), pp.115-152.

Wu, S., Hofman, J.M., Mason, W.A. and Watts, D.J., 2011, March. Who says what to whom on Twitter. In *Proceedings of the 20th international conference on World wide web* (pp. 705-714). ACM.

Ke, X., 2010, January. A Social Networking Services System based on the "Six Degrees of Separation" Theory and Damping Factors. In *Future Networks, 2010. ICFN'10. Second International Conference on* (pp. 438-441). IEEE.

Xuguang, W., Minqing, Z. and Yiliang, H., 2012, July. Research on centrality of node importance in scale-free complex networks. In *Control Conference (CCC), 2012 31st Chinese* (pp. 1073-1077). IEEE.

Xuguang, W., Minqing, Z. and Yiliang, H., 2012, July. Research on centrality of node importance in scale-free complex networks. In *Control Conference (CCC), 2012 31st Chinese* (pp. 1073-1077). IEEE.

YouTube (2016) *Statistics*. Available at: https://www.youtube.com/yt/press/statistics.html (Accessed: 7 November 2016).

Yu, H., 2005, August. SVM selective sampling for ranking with application to data retrieval. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining* (pp. 354-363). ACM.

Zhang, Y. and Leezer, J., 2009, August. Emergence of social norms in complex networks. In *Computational Science and Engineering, 2009. CSE'09. International Conference on* (Vol. 4, pp. 549-555). IEEE.

Zhang, H., Liu, B., Weng, X. and Yu, C., 2012, July. Can online social friends help to improve data swarming performance?. In *2012 21st International Conference on Computer Communications and Networks (ICCCN)* (pp. 1-7). IEEE.

Zhang, S. and Liu, J., 2006, December. From local behaviors to the dynamics in an agent network. In *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence* (pp. 572-580). IEEE Computer Society.

8.1 Appendix

8.1.1 Screen shots, graphs and code of virtual network design

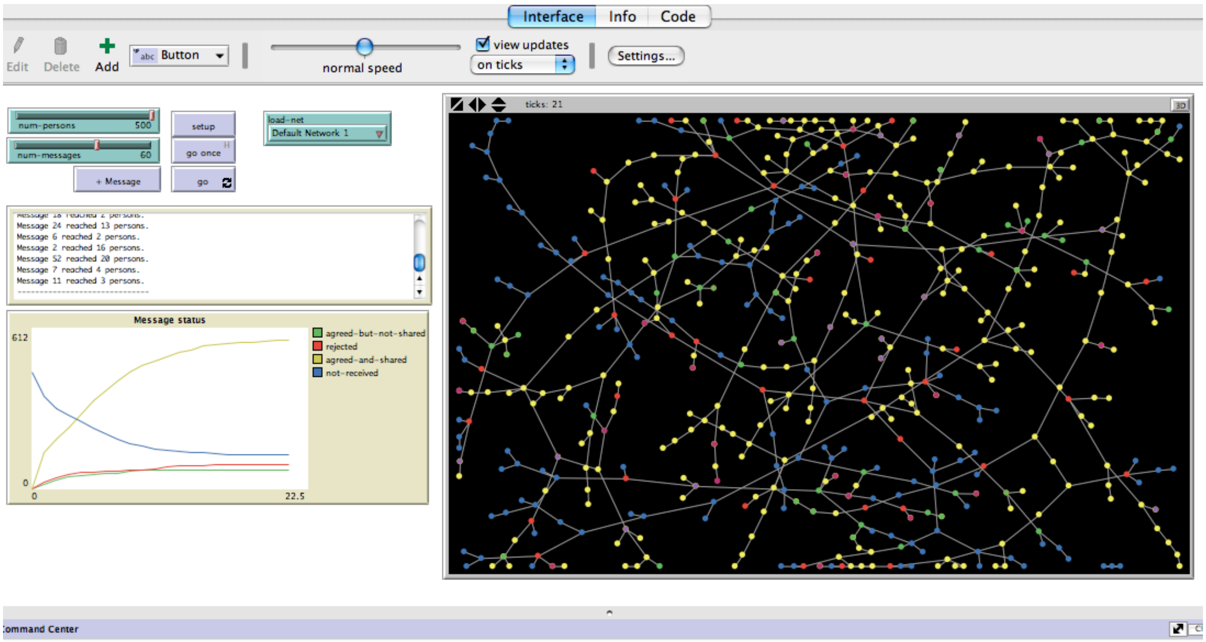

Figure. 8.1 Sample Netlogo Screen Shot; experiment for result 1 showing node interaction with number of messages and number of persons (other nodes) reached.

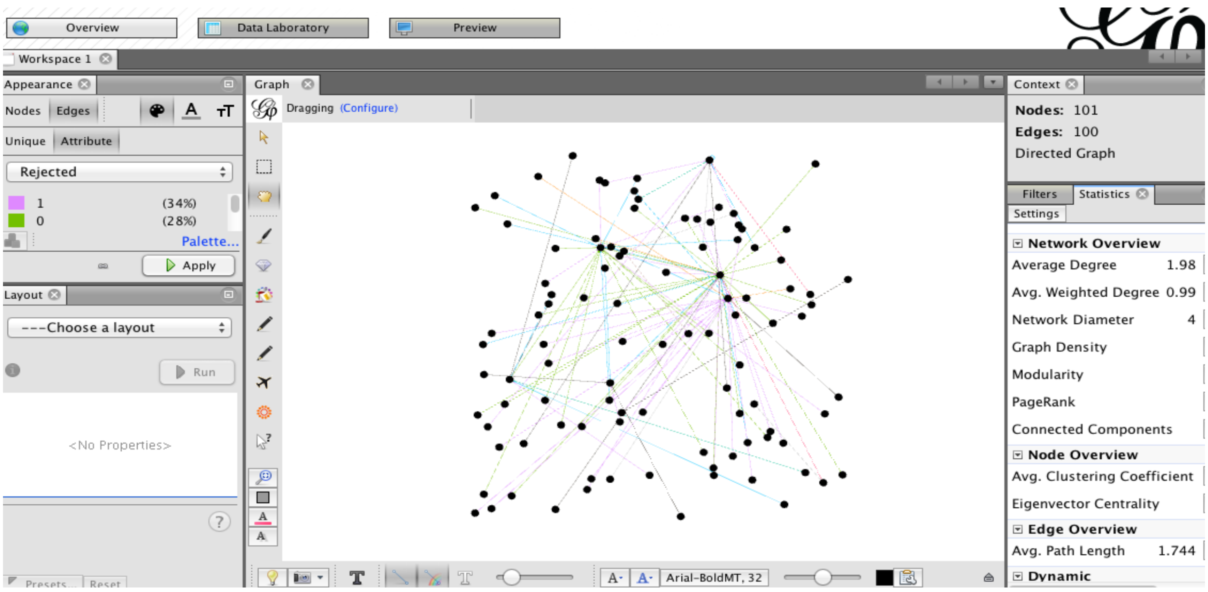

Figure 8.2: Screen shot of graph in Gephi for Results 1 showing the number of nodes and the corresponding number of edges with an average node degree of 1.98.

Results Run 1 table.

Message ID	Accepted	Shared	Rejected
1	0	2	1
2	1	3	0
3	3	44	2
4	2	1	0
5	1	9	2
6	2	12	4
7	0	1	1
8	2	4	1
9	0	5	2
10	6	116	12
11	1	28	6
12	5	50	4
13	1	6	2
14	0	1	1
15	2	18	2
16	0	10	2
17	1	3	0
18	0	4	1
19	2	4	2
20	0	5	4
21	0	3	1
22	0	1	1
23	2	1	0
24	2	9	1
25	0	3	1
26	5	27	4
27	4	17	2
28	1	1	0
29	0	1	1
30	1	1	0
31	1	2	0
32	1	4	0
33	1	14	4
34	3	8	1
35	1	7	2
36	4	6	4
37	0	1	1
38	11	121	9
39	1	1	0
40	2	5	0
41	0	2	1
42	0	4	1
43	1	1	0
44	1	1	0

45	1	1	1
46	1	5	0
47	3	46	4
48	6	49	10
49	1	2	0
50	0	6	3
51	0	3	3
52	0	1	1
53	1	2	1
54	1	1	0
55	1	3	0
56	2	2	0
57	2	1	0
58	2	2	0
59	0	2	1
60	1	1	0
61	3	12	1
62	3	5	1
63	0	11	5
64	2	6	0
65	2	4	1
66	1	5	1
67	2	2	0
68	2	11	2
69	5	9	1
70	0	7	1
71	3	61	8
72	4	17	2
73	0	2	1
74	4	20	4
75	2	17	1
76	1	1	0
77	5	15	2
78	1	2	1
79	1	7	2
80	3	46	6
81	0	4	1
82	2	3	0
83	1	1	0
84	1	7	3
85	1	1	0
86	4	6	1
87	1	6	2
88	0	1	1
89	3	54	7
90	2	9	2

91	4	83	8
92	7	50	4
93	0	1	1
94	0	1	1
95	1	1	0
96	1	7	1
97	3	36	2
98	1	3	1
99	1	1	0
100	3	16	4

Table 8.1: Complete table of result 1 showing 100 messages.

8.2 Netlogo code of virtual network design

```
breed[ persons person ]

breed[ messages message ]

persons-own[ gender age status religion msgaccepted msgrejected msgshared]

messages-own[ id owner topic location nearby parent cacc cshr crej]

globals[list1 list2 idm]


to setup

  clear-all

  set-default-shape persons "circle"

  set-default-shape messages "dot"

  set list1 []
```

```
  set list2 []
  set idm 1
  create-persons num-persons [ set color blue set msgaccepted [] set msgrejected [] set
msgshared []]
  ask persons [ create-link-with one-of other persons ]

  repeat 600 [ layout ]
  ;; leave space around the edges
  ask persons [ setxy 0.98 * xcor 0.98 * ycor ]

  create-ordered-messages num-messages [
    set color one-of [ 13 23 33 43 53 63]
    set id idm
    set size 2
    set location one-of persons
    set owner [who] of location
    set nearby [who] of [link-neighbors] of location
    set topic one-of ["politics" "entertainment" "travel" "science" "lifestyle" "opinion" "health"
"world" "sports" "technology" "movies"]
    set parent [who] of self
    ownerSetup owner [who] of self
    set idm idm + 1
    move-to location
  ]

  initialInfo
  reset-ticks
end

to layout
```

```
  layout-spring persons links 0.2 2 1

end

to go

  foreach list1[

    ask person item 0 ? [

        ask link-neighbors [

          if member? item 1 ? msgaccepted = false and member? item 1 ? msgrejected = false
and member? item 1 ? msgshared = false [

          ifelse (random-float 100 < acceptance-chance) [

            ifelse ((random-float 100) < spread-chance)[

            set color yellow

            set msgshared lput item 1 ? msgshared

            set msgshared remove-duplicates msgshared

            set list2 lput se [who] of self item 1 ? list2]

            [

            set color green

            set msgaccepted lput item 1 ? msgaccepted

            set msgaccepted remove-duplicates msgaccepted]

            ]

            [

            set color red

            set msgrejected lput item 1 ? msgrejected

            set msgrejected remove-duplicates msgrejected]

          ]
```

```
        ]
      ]
    ]

  ask messages[
    countAccepted [who] of self
    countShared [who] of self
    countRejected [who] of self
  ]

  outputInfo

  let temp []
  set temp list1
  set list1 list2
  set list2 temp

  display
  tick
end



to ownerSetup [o m]
    ask person o [
      set color yellow
      set msgaccepted lput [who] of myself msgaccepted
      set list1 lput se [who] of self m list1
       ]
end
```

```
to countAccepted [m]
let caccepted 0
foreach [who] of persons [
  if (member? m [msgaccepted] of person ? = true)[
    set caccepted caccepted + 1
    ]
]
ask message m [ set cacc caccepted ]

end

to countShared [m]
let cshared 0
foreach [who] of persons [
  if (member? m [msgshared] of person ? = true)[
    set cshared cshared + 1
    ]
]
ask message m [ set cshr cshared ]

end

to countRejected [m]
let crejected 0
foreach [who] of persons [
  if (member? m [msgshared] of person ? = true)[
    set crejected crejected + 1
    ]
```

```
]
ask message m [ set crej crejected ]

end

; old go procedure execution
to executeMessage
ask messages [
  foreach nearby [
    if [color] of person ? = blue [
      ifelse ((random-float 100) < acceptance-chance)[

        ask person ? [
         set color green
         set msgaccepted lput [parent] of messages-here msgaccepted
         set msgaccepted remove-duplicates msgaccepted]



       if ((random-float 100) < spread-chance)[
         hatch 1 [
           set location person ?
           set nearby [who] of [link-neighbors] of location
           move-to location ]

         ask person ? [
         set color yellow
         set msgshared lput [parent] of messages-here msgshared
         set msgshared remove-duplicates msgshared]]]
      [
```

```
        ask person ? [
          set color red
          set msgrejected lput [parent] of messages-here msgrejected
          set msgrejected remove-duplicates msgrejected]]
      ]

    ]

  ]

end

; for visualization - not currently being used
to fade
  ask persons with [color != blue]
    [ set color color + 1
      if color > yellow + 4
        [ set color blue ] ]
end

to initialInfo

  output-type "There are " output-type count messages output-type " messages in this network." output-print ""
  foreach [who] of messages[
    output-type "Message " output-type [id] of message ? output-type " with topic " output-type [topic] of message ? output-print ""
  ]

  output-print "------------------------------"
```

```
end

to outputInfo

  foreach [who] of messages[
    output-type "Message " output-type [id] of message ? output-type " reached " output-type [cacc] of message ? + [cshr] of message ?  output-type " persons." output-print " "
  ]
  output-print "------------------------------"
end
```

2. Final Netlogo simulation Code;

```
breed[ persons person ]
breed[ messages message ]
persons-own[ gender age status religion msgaccepted msgrejected msgshared pref conn pref1 pref2 pref3 recall?]
messages-own[ id owner topic info location parent cacc cshr crej nacc nshr nrej]
globals[list1 list2 list3 idm csv fileList topics infodetails]

to setup
  clear-all
  set-default-shape persons "circle"
  set-default-shape messages "dot"
```

```
set list1 []
set list2 []
set list3 []
 set topics ["Any" "Politics" "Entertainment" "Travel" "Science" "Lifestyle" "Health"
"World" "Football" "Technology" "Education" "Rugby" "Tennis" "Swimming" "Golf"
"Skiing" "Fashion" "Music" "Shopping"]
set infodetails ["True" "False" "Neither True Nor False"]
set idm 1
openFile

loadNetwork
loadLinks

  create-ordered-messages num-messages [
    set info one-of [0 1 2]
    set color item info [55 115 125]
    set id idm
    set size 2
    set location one-of persons
    set owner [who] of location
    set topic first [pref] of location
    set parent [who] of self
    ownerSetup owner [who] of self
    set idm idm + 1
    move-to location

]

initialInfo
reset-ticks
```

```
end

to openFile
file-open "pref5000.csv"
let counter 1

while [counter <= num-persons][

set csv file-read-line
set csv word csv ","

 set filelist []

  while [not empty? csv]
   [ let $x position "," csv
     let $item substring csv 0 $x
     carefully [set $item read-from-string $item][]
     set fileList lput $item fileList
     set csv substring csv ($x + 1) length csv

   ]

create-ordered-persons 1 [

  set color blue

  set msgaccepted []

  set msgrejected []
```

```
  set msgshared []

  set pref fileList
  ]

set counter counter + 1
]
file-close
end



to loadNetwork
let netFile ifelse-value (load-net = "Default Network 1") [ "defaultnet1.csv" ]
[ "defaultnet2.csv" ]

file-open netFile
let counter 1

while [counter <= num-persons][

set csv file-read-line
set csv word csv ","

 set filelist []

  while [not empty? csv]
   [ let $x position "," csv
     let $item substring csv 0 $x
     carefully [set $item read-from-string $item][]
     set fileList lput $item fileList
     set csv substring csv ($x + 1) length csv

   ]

if item 0 fileList <  num-persons [
```

```
ask person item 0 fileList [ setxy (item 1 fileList) (item 2 fileList) set conn item 3 fileList]
]

set counter counter + 1


]
file-close
end

to loadLinks

ask persons with [who < num-persons] [
  foreach conn [
    if ( ? < max [who] of persons) [
    create-link-with person ?
    ]
  ]
]

display
end

to-report check [look values]
  report reduce
    [ifelse-value (?2 = look) [?1 + 1] [?1]] (fput 0 values)
end



to go
let x 0
let y sum [cacc] of messages + sum  [cshr] of messages

 set list3 lput y list3

if check last list3 list3 >= 2 [ outputMessage stop]
```

```
foreach list1[

 ask person item 0 ? [

   ask link-neighbors [

    set x [topic] of message item 1 ?

     if member? item 1 ? msgaccepted = false and member? item 1 ? msgrejected = false
and member? item 1 ? msgshared = false [

    ifelse member? (item 1 ?) [pref] of self = true [

      if position item 1 ? [pref] of self = 0 [
      set color yellow
      set msgshared lput item 1 ? msgshared
      set msgshared remove-duplicates msgshared
      set list2 lput se [who] of self item 1 ? list2]

      if position item 1 ? [pref] of self = 2 [
      set color red
      set msgrejected lput item 1 ? msgrejected
      set msgrejected remove-duplicates msgrejected]

      if position item 1 ? [pref] of self = 2 [
      set color green
      set msgaccepted lput item 1 ? msgaccepted
      set msgaccepted remove-duplicates msgaccepted]

   ][
```

```
let xpos [100]

 foreach msgshared [ if [topic] of message ? = x [ set xpos lput 100 xpos ]]
 foreach msgaccepted [ if [topic] of message ? = x [ set xpos lput 50 xpos ]]
 foreach msgrejected [ if [topic] of message ? = x [ set xpos lput 10 xpos ]]


 ask link-neighbors with [ who != [owner] of message item 1 ? ] [

   ifelse recall? = true [

     if (member? x [pref] of self = true) [
       if position x [pref] of self = 0 [ set xpos lput 100 xpos ]
       if position x [pref] of self = 1 [ set xpos lput 50 xpos ]
       if position x [pref] of self = 2 [ set xpos lput 10 xpos ]
     ]

     foreach msgshared [ if [topic] of message ? = x [ set xpos lput 100 xpos ]]
     foreach msgaccepted [ if [topic] of message ? = x [ set xpos lput 50 xpos ]]
     foreach msgrejected [ if [topic] of message ? = x [ set xpos lput 10 xpos ]]

   ] [ set xpos lput one-of [100 50 10] xpos ]

 ]



 let val one-of modes xpos

  if val = 100 [
    set color yellow
    set msgshared lput item 1 ? msgshared
    set msgshared remove-duplicates msgshared
```

```
          set list2 lput se [who] of self item 1 ? list2]

         if val = 50[
          set color green
          set msgaccepted lput item 1 ? msgaccepted
          set msgaccepted remove-duplicates msgaccepted]

         if val = 10 [
          set color red
          set msgrejected lput item 1 ? msgrejected
          set msgrejected remove-duplicates msgrejected]

      ] ;ifelse
     ] ; member
    ] ;libk-neighbor


   ] ;person
 ] ;for

ask messages[
  countAccepted [who] of self
  countShared [who] of self
  countRejected [who] of self
]

outputInfo

let temp []
set temp list1
set list1 list2
set list2 temp

display
tick
```

```
  if ticks mod 2 = 0 [outputPersons]

  memoryUpdate
end

to ownerSetup [o m]
    ask person o [
      set color yellow
      set msgshared lput [who] of myself msgshared
      set list1 lput se [who] of self m list1
       ]
end

to countAccepted [m]
set nacc []
let caccepted 0
foreach [who] of persons [
  if (member? m [msgaccepted] of person ? = true)[
    set caccepted caccepted + 1
    set nacc lput ? nacc
    ]
]
ask message m [ set cacc caccepted ]

end

to countShared [m]
set nshr []
let cshared 0
foreach [who] of persons [
  if (member? m [msgshared] of person ? = true)[
    set cshared cshared + 1
    set nshr lput ? nshr
    ]
```

```
]
ask message m [ set cshr cshared ]

end

to countRejected [m]
set nrej []
let crejected 0
foreach [who] of persons [
  if (member? m [msgrejected] of person ? = true)[
    set crejected crejected + 1
    set nrej lput ? nrej
    ]
]
ask message m [ set crej crejected ]

end

to addMessage
  create-ordered-messages 1 [
      set info one-of [0 1 2]
      set color item info [55 115 125]
      set id idm
      set size 2
      set location one-of persons
      set owner [who] of location
      set topic first [pref] of location
      set parent [who] of self
      ownerSetup owner [who] of self
      set idm idm + 1
      move-to location

      ask patch [xcor] of location [ycor] of location [ ask neighbors [set pcolor 139]]
    ]
end
```

```
to memoryUpdate
  if ticks mod 3 = 0 [
    ask persons[
      set recall? one-of [true false]]
    ]

end

to initialInfo

    output-type "There are " output-type count messages output-type " messages in this network." output-print "" output-print ""
  foreach [who] of messages[
     output-type "Message " output-type [id] of message ? output-type " with topic category " output-type item [topic] of message ? topics output-print "."
  ]

  output-print "------------------------------"

  output-print "Person agents preferences are as follows: "
  foreach sort [who] of persons[
    output-type "Person " output-type [who] of person ? output-type ": "
    foreach [pref] of person ? [ output-type item ? topics output-type " " ]
     output-print "."
  ]
   output-print "------------------------------"
end

to outputInfo

  foreach [who] of messages[
     output-type "Message " output-type [id] of message ? output-type " reached " output-type [cacc] of message ? + [cshr] of message ?  output-type " persons." output-print " "
  ]
  output-print "------------------------------"
```

```
end

;; Writes Persons info to data.csv
to outputPersons

carefully [

file-open "data.csv"

file-type "Person" file-type "," file-type "Pref1" file-type "," file-type "Pref2" file-type "," file-type "Pref3" file-type "," file-type "Tick" file-type "," file-type "Time" file-print ""
foreach sort-on [who] persons[
  ask ?[

    set pref1 []
    set pref2 []
    set pref3 []
    file-type ? file-type ","

     ;if empty? msgshared = true and empty? msgaccepted = true and empty? msgrejected = true [ stop ]

       if not empty? msgshared = true [foreach msgshared [set pref1 lput item [topic] of message ? topics pref1 ]  ]
      set pref1 lput item (item 0 pref) topics pref1
      set pref1 modes pref1
      file-type modes pref1 file-type ","

      if not empty? msgaccepted = true [foreach msgaccepted [set pref2 lput item [topic] of message ? topics pref2 ] ]
      set pref2 lput item (item 1 pref) topics pref2
      set pref2 modes pref2
      file-type modes pref2 file-type ","

      if not empty? msgrejected = true [foreach msgrejected [set pref3 lput item [topic] of message ? topics pref3 ] ]
```

```
    set pref3 lput item (item 2 pref) topics pref3
    set pref3 modes pref3
    file-type modes pref3 file-type ","

  ]

  file-type ticks file-type ","
  file-type date-and-time
  file-print ""
]

file-close

] [ print error-message ]

end

;; Appends Messages info to data.csv
to outputMessage

carefully [

file-open "data.csv"

file-type "Date Saved: "file-print date-and-time
file-type "Message ID"  file-type "," file-type "Info" file-type "," file-type "Topic" file-type "," file-type "Accepted" file-type "," file-type "Shared" file-type "," file-type "Rejected" file-type "," file-type "Person Agents Who Accepted" file-type "," file-type "Person Agents Who Shared" file-type "," file-type "Person Agents Who Rejected" file-type "," file-type "info" file-print ""
foreach sort-on [id] messages[
  ask ?[
    file-type id file-type ","

    ;if length nacc = 0 and length nshr = 0 and length nrej = 0 [ stop ]
    file-type item [info] of ? infodetails file-type ","
```

```
    file-type item [topic] of ? topics file-type ","
    file-type cacc file-type ","
    file-type cshr file-type ","
    file-type crej file-type ","
    file-type nacc file-type ","
    file-type nshr file-type ","
    file-type nrej file-type ","

  ]
  file-print ""
]

file-close

] [ print error-message ]
end
```